HAL
open science

A MULTI-STEP RICHARDSON-ROMBERG EXTRAPOLATION METHOD FOR STOCHASTIC APPROXIMATION

Noufel Frikha, Lorick Huang

To cite this version:

Noufel Frikha, Lorick Huang. A MULTI-STEP RICHARDSON-ROMBERG EXTRAPOLATION METHOD FOR STOCHASTIC APPROXIMATION. 2014. hal-01064536v1

HAL Id: hal-01064536

https://hal.science/hal-01064536v1

Preprint submitted on 16 Sep 2014 (v1), last revised 9 Mar 2015 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A MULTI-STEP RICHARDSON-ROMBERG EXTRAPOLATION METHOD FOR STOCHASTIC APPROXIMATION

N. Frikha[1] and L. Huang[2]

Abstract. We obtain an expansion of the implicit weak discretization error for the target of stochastic approximation algorithms introduced and studied in [Fri13]. This allows us to extend and develop the Richardson-Romberg extrapolation method for Monte Carlo linear estimator (introduced in [TT90] and deeply studied in [Pag07]) to the framework of stochastic optimization by means of stochastic approximation algorithm. We notably apply the method to the estimation of the quantile of diffusion processes. Numerical results confirm the theoretical analysis and show a significant reduction in the initial computational cost.

1991 Mathematics Subject Classification. 60H35,65C30,65C05.

September 16, 2014.

1. Statement of the Problem

The aim of this paper is to combine a multistep Richardson-Romberg extrapolation method with stochastic approximation (SA) algorithms which are recursive simulation based procedures commonly used in the framework of stochastic optimization. Introduced by Robbins and Monro [RM51], SA algorithms aims at computing a zero of a continuous function $h : \mathbb{R}^d \to \mathbb{R}^d$ which is unknown to the experimenter but can only be estimated through experiments. In this general context, the function h writes $h(\theta) := \mathbb{E}[H(\theta, U)]$ where $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ and U is a $\mathbb{R}^q$-valued random vector. To estimate a zero of h, one devises the following recursive algorithm

$$\theta_{p+1} = \theta_p - \gamma_{p+1} H(\theta_p, U^{p+1}),\ p \geq 0 \tag{1.1}$$

where $(U^p)_{p\geq 1}$ is an *i.i.d.* sequence of random variables with the same law as U defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, θ_0 is independent of the innovation of the algorithm with $\mathbb{E}[|\theta_0|^2] < +\infty$ and $\gamma = (\gamma_p)_{p\geq 1}$ is a deterministic and decreasing sequence of non-negative steps satisfying the usual assumption

$$\sum_{p\geq 1} \gamma_p = +\infty, \quad \text{and} \quad \sum_{p\geq 1} \gamma_p^2 < +\infty. \tag{1.2}$$

When the function h is the gradient of a convex potential, the recursive procedure (1.1) is a stochastic gradient algorithm. Indeed replacing $H(\theta_p, U^{p+1})$ by $h(\theta_p)$ in (1.1) leads to the usual deterministic descent gradient procedure.

Keywords and phrases: Euler scheme, weak error, Richardson-Romberg extrapolation, stochastic approximation algorithm

[1] LPMA, Université Paris Diderot, 5 rue Thomas Mann 75013 Paris, frikha@math.univ-paris-diderot.fr

[2] LPMA, Université Paris Diderot, 5 rue Thomas Mann 75013 Paris, huang@math.univ-paris-diderot.fr

In many applications, notably in computational finance, the sequence of random vectors $(U^p)_{p\geq 1}$ is not directly simulatable (at a reasonable cost) and can only be approximated by another sequence of easily simulatable random vectors $((U^n)^p)_{p\geq 1}$, $n>0$, where U^n (weakly or strongly) approximates U as $n\to+\infty$ with a standard weak discretization error (or bias) $\mathbb{E}[f(U^n)]-\mathbb{E}[f(U)]$ that can be expanded in powers of $n^{-\alpha}$, $\alpha>0$, for a specific class of functions $f\in\mathcal{C}$. One typical situation is when $U=X_T$, $X:=(X_t)_{t\in[0,T]}$ being a q-dimensional diffusion process solution of a stochastic differential equation (SDE) and $U^n=X_T^n$ where $X^n:=(X_t^n)_{t\in[0,T]}$ stands for its standard Euler-Maruyama discretization scheme with time step $\Delta=T/n$, $n\in\mathbb{N}^*$.

Since we are interested in the computation of the zero θ^* of h given by $h(\theta):=\mathbb{E}[H(\theta,U)]$ where $H:\mathbb{R}^d\times\mathbb{R}^q\to\mathbb{R}^d$ and the function h is generally neither known nor computable since the random variable U cannot be easily simulated, estimating θ^* by devising directly the recursive scheme (1.1) is not possible. Therefore, two steps are needed to compute θ^*:

- the first step consists in approximating the zero θ^* of h by the zero $\theta^{*,n}$ of the function h^n defined by $h^n(\theta):=\mathbb{E}[H(\theta,U^n)]$, $\theta\in\mathbb{R}^d$. It induces *an implicit discretization error* which writes

$$\mathcal{E}_D(n,T,b,\sigma,H):=\theta^*-\theta^{*,n}.$$

Under mild assumptions on h and h^n, it is proved in [Fri13] that $\theta^{*,n}$ converges to θ^* as n goes to infinity. Moreover, if the *standard weak discretization error* is of order $n^{-\alpha}$, $\alpha\in(0,1)$, that is $\forall\theta\in\mathbb{R}^d,\ h^n(\theta)-h(\theta)=\Lambda_1^0(\theta)n^{-\alpha}+o(n^{-\alpha})$, with $\Lambda_1^0:\mathbb{R}^d\to\mathbb{R}^d$, then (under additional mild assumptions) this rate of convergence transfers to the *implicit discretization error* that is $\mathcal{E}_D(n,T,b,\sigma,H)=\Theta_1 n^{-\alpha}+o(n^{-\alpha})$ for some $\Theta_1\in\mathbb{R}^d$.

- the second step consists in approximating $\theta^{*,n}$ using $M\in\mathbb{N}^*$ steps of the following SA scheme

$$\theta_{p+1}^n=\theta_p^n-\gamma_{p+1}H(\theta_p^n,(U^n)^{p+1}),\ p\in[\![0,M-1]\!], \tag{1.3}$$

where $((U^n)^p)_{p\in[\![1,M]\!]}$ is an i.i.d. sequence of random variables with the same law as U^n, θ_0^n is independent of the innovation of the algorithm with $\sup_{n\geq 1}\mathbb{E}|\theta_0^n|^2<+\infty$ and $\gamma=(\gamma_p)_{p\geq 1}$ is a sequence of non-negative deterministic and decreasing steps satisfying (1.2). This induces a *statistical error* which writes

$$\mathcal{E}_S(n,M,\gamma,T,H):=\theta^{*,n}-\theta_M^n.$$

Regarding the *statistical error*, it is well-known that under mild assumptions the Robbins-Monro theorem guarantees that for each $n\in\mathbb{N}^*$, $\lim_{M\to+\infty}\mathcal{E}_S(n,M,\gamma,T,H)=0$. Moreover, under additional technical assumptions, a central limit theorem (CLT) holds at rate $\gamma^{-1/2}(M)$ that is $\gamma^{-1/2}(M)\mathcal{E}_S(n,M,\gamma,T,H)$ converges in distribution to a normally distributed random variable. The reader may also refer to [FM12] and [FF13] for some recent developments on non-asymptotic deviation bounds for the statistical error.

The global error between θ^*, the quantity to estimate, and its implementable approximation θ_M^n can be decomposed as follows:

$$\begin{aligned}\mathcal{E}_{glob}(M,\gamma,H)&=\theta^*-\theta^{*,n}+\theta^{*,n}-\theta_M^n\\&:=\mathcal{E}_D(n,T,b,\sigma,H)+\mathcal{E}_S(n,M,\gamma,T,H).\end{aligned}$$

The first aim of this paper is to prove the existence of an expansion for the *implicit discretization error*, that is, under mild assumptions (see Section 2) on h and h^n, $\mathcal{E}_D(n,T,b,\sigma,H)$ can be expanded as follows

$$\forall R\in\mathbb{N}^*,\quad \theta^{*,n}-\theta^*=\frac{C_1}{n^\alpha}+\cdots+\frac{C_R}{n^{\alpha R}}+o\left(\frac{1}{n^{\alpha R}}\right) \tag{1.4}$$

where $(C_1,\cdots,C_R)\in(\mathbb{R}^d)^R$. Then taking advantage of (1.4) we devise a multistep Richardson-Romberg extrapolation method for stochastic optimization by means of stochastic approximation algorithm. The principle of Richardson-Romberg extrapolation is to reduce the bias produced by the *implicit discretization error* by

combining two estimators with different step size. To be more precise, one considers the Richardson-Romberg SA estimator

$$\Theta_M^{n,2n} = 2^\alpha \theta_M^{2n} - \theta_M^n$$

where $(\theta_M^{2n}, \theta_M^n)$ is obtained using M steps of two SA schemes devised with the i.i.d. sequence $((U^{2n}, U^n)^p)_{p\in[\![1,M]\!]}$ of random variables with the same law as (U^{2n}, U^n). Under standard assumptions, this linear combination of SA estimators *a.s.* converges to the target $2^\alpha \theta^{*,2n} - \theta^{*,n}$ as the number of steps M goes to infinity. The key observation is that this new target satisfies the following implicit error expansion of order 2

$$2^\alpha \theta^{*,2n} - \theta^{*,n} - \theta^* = \left(\frac{1}{2^{2\alpha-1}} - 1\right)\frac{C_2}{n^{2\alpha}} + o\left(\frac{1}{n^{2\alpha}}\right).$$

Moreover, in the spirit of [Pag07], we show how to control the asymptotic $L^1(\mathbb{P})$-norm of the distance between the new estimator $\Theta_M^{n,2n}$ and its target $2^\alpha \theta^{*,2n} - \theta^{*,n}$ as n goes to infinity. Then, it is natural to iterate this extrapolation to obtain a new stochastic approximation estimator with an implicit discretization error of order $n^{-\alpha R}$ for any $R \in \mathbb{N}^*$. This extension called multi-step Richardson-Romberg extrapolation is deeply investigated in [Pag07] for Monte Carlo linear estimator in the framework of discretization of diffusion processes.

The aim of this paper is to investigate the Richardson-Romberg SA method. Our purpose is to show that the principle of multi-step Richardson-Romberg extrapolation for Monte Carlo linear estimator can be extended to the framework of stochastic optimization by means of SA algorithm. We notably prove that the new estimator outperforms the standard SA estimator in terms of computational cost.

The paper is organized as follows: in Section 2 we provide an expansion of the implicit discretization error in powers of $n^{-\alpha}$ under mild assumptions. Then we take advantage of this expansion to propose a multi-step Richardson-Romberg method by means of SA. In Section 3 is presented an illustration of the method to the estimation of the quantile of a stochastic differential equation (SDE) driven by a stable process. In Section 4 numerical results are carried out to confirm the theoretical analysis. Finally, Section 5 is devoted to theoretical results which are useful throughout the paper.

2. Main results

This section is divided in two parts. In the first one we obtain a general result concerning the expansion of the implicit discretization error. In the second one, we take advantage of this result to develop a Richardson-Romberg extrapolation method for stochastic optimization by means of SA algorithms.

2.1. Expansion of the implicit discretization error

We first provide a result concerning the convergence of the sequence $(\theta^{*,n})_{n\geq 1}$ towards θ^*. For a proof the reader may refer to [Fri13].

Proposition 2.1. *For all $n \in \mathbb{N}^*$, assume that h and h^n satisfy the mean reverting assumption:*

$$\forall \theta \neq \theta^*,\ \langle \theta - \theta^*, h(\theta)\rangle > 0 \quad \text{and} \quad \forall \theta \neq \theta^{*,n},\ \langle \theta - \theta^{*,n}, h^n(\theta)\rangle > 0.$$

Moreover, suppose that $(h^n)_{n\geq 1}$ converges locally uniformly towards h. Then, one has

$$\theta^{*,n} \to \theta^* \quad \text{as} \quad n \to +\infty.$$

Here we will investigate an expansion of the error term $\theta^{*,n} - \theta^*$ in powers of $n^{-\alpha}$. Through the document, we will refer to **[H-k]** the following set of assumptions:

(1) For all $\theta \in \mathbb{R}^d$,

$$h(\theta) - h^n(\theta) = \frac{\Lambda_1^0(\theta)}{n^\alpha} + \cdots + \frac{\Lambda_k^0(\theta)}{n^{\alpha k}} + o\left(\frac{1}{n^{\alpha k}}\right). \tag{2.5}$$

(2) $h, h^n \in \mathcal{C}^k(\mathbb{R}^d, \mathbb{R}^d)$ and for all $l \leq k-1$, for all $\theta \in \mathbb{R}^d$,

$$D^l h^n(\theta) - D^l h(\theta) = \frac{\Lambda_1^l(\theta)}{n^\alpha} + \cdots + \frac{\Lambda_{k-l}^l(\theta)}{n^{\alpha(k-l)}} + o\left(\frac{1}{n^{\alpha(k-l)}}\right) \tag{2.6}$$

(3) For all $l \in [\![1, k]\!]$, $(D^l h^n)_{n\geq 1}$ converges locally uniformly towards $D^l h$.
(4) $Dh(\theta^*)$ is invertible.

Proposition 2.2. *Assume that $\theta^{*,n} \to \theta^*$ as $n \to +\infty$. Under **[H-1]**, one has*

$$n^\alpha\left(\theta^{*,n} - \theta^*\right) \underset{n\to\infty}{\longrightarrow} Dh(\theta^*)^{-1}\Lambda_1^0(\theta^*). \tag{2.7}$$

Proof. Observe that one has $h^n(\theta^{*,n}) - h^n(\theta^*) = -h^n(\theta^*) = h(\theta^*) - h^n(\theta^*)$. On the one hand, writing Taylor's formula with integral remainder yields:

$$h^n(\theta^{*,n}) - h^n(\theta^*) = \int_0^1 dt Dh^n(t\theta^{*,n} + (1-t)\theta^*)(\theta^{*,n} - \theta^*). \tag{2.8}$$

On the other hand, from the discretization error, we have $h(\theta^*) - h^n(\theta^*) = \Lambda_1^0(\theta^*)n^{-\alpha} + o\left(n^{-\alpha}\right)$.

Since $\theta^{*,n} \underset{n\to\infty}{\longrightarrow} \theta^*$, $Dh(\theta^*)$ is invertible, and $(Dh^n)_{n\geq 1}$ converges uniformly locally to Dh, for n large enough, the matrix $\int_0^1 Dh^n(t\theta^{*,n} + (1-t)\theta^*)dt$ is invertible. Multiplying both sides of (2.8) by $n^{-\alpha}$ finally yields

$$n^\alpha\left(\theta^{*,n} - \theta^*\right) = \left(\int_0^1 Dh^n(t\theta^{*,n} + (1-t)\theta^*)dt\right)^{-1}\left(\Lambda_1^0(\theta^*) + o(1)\right) \underset{n\to\infty}{\longrightarrow} Dh(\theta^*)^{-1}\Lambda_1^0(\theta^*).$$

□

Let us note that Proposition 2.2 provides a first order expansion of $\theta^{*,n} - \theta^*$, that is $\theta^{*,n} - \theta^* = C_1 n^{-\alpha} + o(n^{-\alpha})$. We now give a generalization of this first result.

Theorem 2.1. *Assume that $\theta^{*,n} \to \theta^*$, $n \to +\infty$, and that **[H-p]** holds for some $p \in \mathbb{N}^*$. Then, $\theta^{*,n} - \theta^*$ has an expansion up to order p, that is, the following expansion holds:*

$$\theta^{*,n} - \theta^* = \frac{C_1}{n^\alpha} + \cdots + \frac{C_p}{n^{\alpha p}} + o\left(\frac{1}{n^{\alpha p}}\right).$$

Proof. If **[H-p]**, $p \in \mathbb{N}^*$, holds then Proposition 2.2 gives a first order expansion for $\theta^{*,n} - \theta^*$. We now prove the inductive step that is if $\theta^{*,n} - \theta^*$ has an expansion of order $k-1$ then an expansion exists at order k, for $k \leq p$. The basic idea does not change from the previous computation. From the development of the discretization error, we have:

$$h(\theta^*) - h^n(\theta^*) = \frac{\Lambda_1^0(\theta^*)}{n^\alpha} + \cdots + \frac{\Lambda_k^0(\theta^*)}{n^{\alpha k}} + o\left(\frac{1}{n^{\alpha k}}\right). \tag{2.9}$$

On the other hand, we write a Taylor's expansion of h^n up to the same order $k-1$:

$$h^n(\theta^{*,n}) - h^n(\theta^*) = Dh^n(\theta^*)(\theta^{*,n} - \theta^*) + \cdots + \frac{1}{(k-1)!}D^{k-1}h^n(\theta^*)(\theta^{*,n} - \theta^*)^{(k-1)} + R_{k-1}^n(\theta^{*,n} - \theta^*), \tag{2.10}$$

with the remainder in integral form satisfying:

$$R_{k-1}^n(\theta^{*,n} - \theta^*) = \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!}D^k h^n(t\theta^{*,n} + (1-t)\theta^*)(\theta^{*,n} - \theta^*)^{(k)}dt = \frac{1}{k!}D^k h(\theta^*)(\theta^{*,n} - \theta^*)^{(k)} + o\left(\frac{1}{n^{\alpha k}}\right)$$

where we used that $(D^k h^n)_{n\geq 1}$ converges locally uniformly to $D^k h$, $k \in [\![1,p]\!]$, and $\theta^{*,n} - \theta^* = \mathcal{O}(n^{-\alpha})$ for the last equality. The expansions (2.6) allow us to replace the derivatives of h^n by the derivatives of h in (2.10) at the cost of an error term, that is:

$$\begin{aligned} h^n(\theta^{*,n}) - h^n(\theta^*) = Dh(\theta^*)(\theta^{*,n} - \theta^*) &+ \left(\frac{\Lambda_1^1(\theta^*)}{n^\alpha} + \cdots + \frac{\Lambda_{k-1}^1(\theta^*)}{n^{\alpha(k-1)}} + o\left(\frac{1}{n^{\alpha(k-1)}}\right)\right)(\theta^{*,n} - \theta^*) \\ &+ \cdots + \frac{1}{(k-1)!}\left(D^{k-1}h(\theta^*) + \frac{\Lambda_1^{k-1}(\theta^*)}{n^\alpha} + o\left(\frac{1}{n^\alpha}\right)\right)(\theta^{*,n} - \theta^*)^{(k-1)} \\ &+ \frac{1}{k!}D^k h(\theta^*)(\theta^{*,n} - \theta^*)^{(k)} + o\left(\frac{1}{n^{\alpha k}}\right). \end{aligned}$$

Since $h^n(\theta^{*,n}) - h^n(\theta^*) = -h^n(\theta^*) = h(\theta^*) - h^n(\theta^*)$ and $Dh(\theta^*)$ is invertible, the previous equality implies

$$\begin{aligned} &\frac{Dh(\theta^*)^{-1}\Lambda_1^0(\theta^*)}{n^\alpha} + \cdots + \frac{Dh(\theta^*)^{-1}\Lambda_k^0(\theta^*)}{n^{\alpha k}} + o\left(\frac{1}{n^{\alpha k}}\right) = \\ &\theta^{*,n} - \theta^* + \left(\frac{Dh(\theta^*)^{-1}\Lambda_1^1(\theta^*)}{n^\alpha} + \cdots + \frac{Dh(\theta^*)^{-1}\Lambda_{k-1}^1(\theta^*)}{n^{\alpha(k-1)}} + o\left(\frac{1}{n^{\alpha(k-1)}}\right)\right)(\theta^{*,n} - \theta^*) \\ &+ \cdots + \frac{1}{(k-1)!}\left(Dh(\theta^*)^{-1}D^{k-1}h(\theta^*) + \frac{Dh(\theta^*)^{-1}\Lambda_1^{k-1}(\theta^*)}{n^\alpha} + o\left(\frac{1}{n^\alpha}\right)\right)(\theta^{*,n} - \theta^*)^{(k-1)} \\ &+ \frac{1}{k!}Dh(\theta^*)^{-1}D^k h(\theta^*)(\theta^{*,n} - \theta^*)^{(k)} + o\left(\frac{1}{n^{\alpha k}}\right). \end{aligned}$$

The last equation should be seen as a "bootstrap" for $\theta^{*,n} - \theta^*$, that is:

$$\begin{aligned} \theta^{*,n} - \theta^* = &\frac{Dh(\theta^*)^{-1}\Lambda_1^0(\theta^*)}{n^\alpha} + \cdots + \frac{Dh(\theta^*)^{-1}\Lambda_k^0(\theta^*)}{n^{\alpha k}} + o\left(\frac{1}{n^{\alpha k}}\right) \\ &- \left(\frac{Dh(\theta^*)^{-1}\Lambda_1^1(\theta^*)}{n^\alpha} + \cdots + \frac{Dh(\theta^*)^{-1}\Lambda_{k-1}^1(\theta^*)}{n^{\alpha(k-1)}} + o\left(\frac{1}{n^{\alpha(k-1)}}\right)\right)(\theta^{*,n} - \theta^*) \\ &- \cdots \\ &- \frac{Dh(\theta^*)^{-1}}{(k-1)!}\left(D^k h(\theta^*) + \frac{\Lambda_1^{k-1}(\theta^*)}{n^\alpha} + o\left(\frac{1}{n^\alpha}\right)\right)(\theta^{*,n} - \theta^*)^{(k-1)} \\ &- \frac{1}{k!}Dh(\theta^*)^{-1}D^k h(\theta^*)(\theta^{*,n} - \theta^*)^{(k)} + o\left(\frac{1}{n^{\alpha k}}\right), \end{aligned} \tag{2.11}$$

The idea now is to plug the expansion of $\theta^{*,n} - \theta^*$ in the right hand side of (2.11) and check that the first remainder term comes at order $o(n^{-\alpha k})$. It is clear that on the first line the remainder term is of order $o(n^{-\alpha k})$. Note that for $l \in [\![1,k]\!]$, $D^l h(\theta^*)$, $\Lambda_1^l(\theta^*)$, $\ldots$, $\Lambda_{k-l}^l(\theta^*)$ and $o\left(n^{-\alpha(k-l)}\right)$, appearing in (2.11) are multilinear maps from $(\mathbb{R}^d)^l$ to $\mathbb{R}^d$. Thus, for any $l \in [\![2,k]\!]$, the generic l-th term writes in the i-th component:

$$\begin{aligned}
&\left(\left(D^l h(\theta^*) + \frac{\Lambda_1^l(\theta^*)}{n^\alpha} + \cdots + \frac{\Lambda_{k-l}^l(\theta^*)}{n^{\alpha(k-l)}} + o\left(\frac{1}{n^{\alpha(k-l)}}\right)\right)(\theta^{*,n} - \theta^*)^{(l)}\right)_i \\
&= \sum_{i_1+\cdots+i_d=l} \frac{1}{i_1!\cdots i_d!}\Lambda_{i_1,\cdots,i_d}(\theta^{*,n} - \theta^*)_1^{i_1} \times \cdots \times (\theta^{*,n} - \theta^*)_d^{i_d} + o\left(\frac{1}{n^{\alpha(k-l)}}\right)(\theta^{*,n} - \theta^*)_1^{i_1} \times \cdots \times (\theta^{*,n} - \theta^*)_d^{i_d} \\
&= \sum_{i_1+\cdots+i_d=l} \frac{1}{i_1!\cdots i_d!}\Lambda_{i_1,\cdots,i_d}(\theta^{*,n} - \theta^*)_1^{i_1} \times \cdots \times (\theta^{*,n} - \theta^*)_d^{i_d} + o\left(\frac{1}{n^{\alpha k}}\right) \qquad (2.12)
\end{aligned}$$

where $\Lambda_{i_1,\cdots,i_d} = \frac{\partial^l h_i}{\partial\theta_1^{i_1}\cdots\partial\theta_d^{i_d}}(\theta^*) + \frac{(\Lambda_1^l(\theta^*))_i}{n^\alpha} + \cdots + \frac{(\Lambda_{k-l}^l(\theta^*))_i}{n^{\alpha(k-l)}}$ with $(\Lambda_j^l(\theta^*))_i$ for $j \in [\![1, k-l]\!]$ satisfying $\frac{\partial^l h_i}{\partial\theta_1^{i_1}\cdots\partial\theta_d^{i_d}}(\theta^*) - \frac{\partial^l h_i^n}{\partial\theta_1^{i_1}\cdots\partial\theta_d^{i_d}}(\theta^*) = (\Lambda_1^l(\theta^*))_i/n^\alpha + \cdots + (\Lambda_{k-l}^l(\theta^*))_i/n^{\alpha(k-l)} + o(1/n^{\alpha(k-l)})$ and where we used that $(\theta^{*,n} - \theta^*)_1^{i_1} \times \cdots \times (\theta^{*,n} - \theta^*)_d^{i_d} = \mathcal{O}(1/n^{\alpha l})$ for the last equality. Now, replacing $(\theta^{*,n} - \theta^*)_i$ by its expansion, we observe that the generic term in (2.12) satisfies

$$\begin{aligned}
&\Lambda_{i_1,\cdots,i_d}\left(\frac{C_1^1}{n^\alpha} + \cdots + \frac{C_{k-1}^1}{n^{\alpha(k-1)}} + o\left(\frac{1}{n^{\alpha(k-1)}}\right)\right)^{i_1} \times \cdots \times \left(\frac{C_1^d}{n^\alpha} + \cdots + \frac{C_{k-1}^d}{n^{\alpha(k-1)}} + o\left(\frac{1}{n^{\alpha(k-1)}}\right)\right)^{i_d} \\
&= \Lambda_{i_1,\cdots,i_d}\left(\frac{\tilde{C}}{n^{\alpha l}} + \cdots + o\left(\frac{1}{n^{\alpha(k+l-2)}}\right)\right) \\
&= \Lambda_{i_1,\cdots,i_d}\left(\frac{\tilde{C}}{n^{\alpha l}} + \cdots + o\left(\frac{1}{n^{\alpha k}}\right)\right)
\end{aligned}$$

where $\tilde{C} = (C_1^1)^{i_1} \times \cdots \times (C_1^d)^{i_d}$. We clearly see that the expression above yields an expansion in powers of $n^{-\alpha}$ with a remainder at order $o(n^{-\alpha k})$. Formally, as the power in the expansion (2.6) goes down, the power in the derivatives grows, compensating exactly and giving the right order in the remainder.

Finally, we expand the previous equation and group together the different terms with respect to the power of $n^{-\alpha}$. As we observed above, the remainder term is at order $o\left(n^{-\alpha k}\right)$, because of the compensation between the power in the expansion (2.6) and the order of the Taylor expansion. This completes the proof. □

2.2. **Multi-step Richardson-Romberg extrapolation for stochastic approximation**

Multi-step Richardson-Romberg extrapolation was successfully applied in the context of Monte Carlo linear estimator for the computation of $\mathbb{E}[f(X_T)]$, where $f : \mathbb{R}^d \to \mathbb{R}$ (with possible extension to the case of path-dependent options) and X is the (unique) strong solution to a SDE, see [Pag07]. In this section, we propose a multi-step Richardson-Romberg SA estimator with a control of the statistical error. We proceed as follows. Let $R \geq 2$ be an integer. To devise a SA estimator whose target has an implicit discretization error of order $n^{-\alpha R}$ as $n \to +\infty$, we introduce a sequence of R random vectors $\{U^{rn}, r \in [\![1, R]\!]\}$, $n \in \mathbb{N}^*$. Throughout this section we will assume that this sequence satisfies $U^{rn} \xrightarrow{\mathbb{P}} U^r$ as $n \to +\infty$ with $U^r \overset{d}{=} U$, $r \in [\![1, R]\!]$, all variables being defined on the same probability space. If assumption [**H-R**] holds then for all $r \in [\![1, R]\!]$ one gets

$$\theta^{*,rn} = \theta^* + \sum_{p=1}^{R-1} \frac{C_p}{r^{\alpha p}}\frac{1}{n^{\alpha p}} + \frac{C_R}{r^{\alpha R}}\frac{1}{n^{\alpha R}}\left(1 + \epsilon_r(n)\right)$$

with $\epsilon_r(n) \to 0$ as $n \to +\infty$. Then, one defines the Vandermonde $Rd \times (R-1)d$ matrix

$$V = \left[\frac{I_d}{r^{\alpha p}}\right]_{1\le r\le R, 1\le p\le R-1}$$

and the extended $Rd \times d$ unit matrix $\mathbf{I} = (I_d, \cdots, I_d)^T$ where I_d is the identity matrix of dimension d. Now we write

$$\begin{pmatrix} \vdots \\ \theta^{*,rn} \\ \vdots \end{pmatrix}_{1\le r\le R} = \mathbf{I}\theta^* + V \begin{pmatrix} \vdots \\ \frac{C_r}{n^{\alpha r}} \\ \vdots \end{pmatrix}_{1\le r\le R-1} + \begin{pmatrix} \vdots \\ \frac{C_R}{r^{\alpha R}} \frac{1}{n^{\alpha R}} \left(1+\epsilon_r(n)\right) \\ \vdots \end{pmatrix}_{1\le r\le R}. \tag{2.13}$$

We consider the $Rd \times d$ weight matrix $\mathbf{w} = (\mathbf{w}_1, \cdots, \mathbf{w}_R)^T$, $\mathbf{w}_i$ being a $d \times d$ matrix for $i \in [\![1, R]\!]$ satisfying

$$\mathbf{w}^T\mathbf{I} = I_d \quad \text{and} \quad \mathbf{w}^T V = 0_{d\times d(R-1)} \tag{2.14}$$

which is equivalent to

$$\tilde{V}\mathbf{w} = E_1 \tag{2.15}$$

with $E_1 = (I_d, 0_{d\times d(R-1)})^T$ and $\tilde{V}$ is the Vandermonde matrix defined by

$$\tilde{V} = \begin{pmatrix} I_d & I_d & \cdots & I_d \\ I_d & \frac{I_d}{2^\alpha} & \cdots & \frac{I_d}{R^\alpha} \\ \vdots & \vdots & \cdots & \vdots \\ I_d & \frac{I_d}{2^{(R-1)\alpha}} & \cdots & \frac{I_d}{R^{(R-1)\alpha}} \end{pmatrix}.$$

Thanks to Cramer's rule, the solution $\mathbf{w}$ to (2.15) is explicitly given by

$$\forall r \in \{1, \cdots, R\}, \quad \mathbf{w}_r = (-1)^{R-r} \frac{r^{\alpha R}}{\prod_{j=0}^{r-1}(r^\alpha - j^\alpha)\prod_{j=r+1}^{R}(j^\alpha - r^\alpha)} I_d. \tag{2.16}$$

Let us note that when $\alpha = 1$ this last expression simplifies to $\mathbf{w}_r = (-1)^{R-r}(r^R/(r!(R-r)!))I_d$, $r = 1, \cdots, R$. The first condition in (2.14) reads $\sum_{r=1}^R \mathbf{w}_r = I_d$ which implies that $\lim_{n\to+\infty} \sum_{r=1}^R \mathbf{w}_r\theta^{*,rn} = \sum_{r=1}^R \mathbf{w}_r\theta^* = \theta^*$. Moreover, multiplying (2.13) on the left by $\mathbf{w}^T$ yields

$$\sum_{r=1}^R \mathbf{w}_r\theta^{*,rn} = \theta^* + C_R \frac{1}{n^{\alpha R}} \tilde{\mathbf{w}}_{R+1} \left(1 + \epsilon_{R+1}(n)\right) \tag{2.17}$$

where

$$\tilde{\mathbf{w}}_{R+1} = \sum_{r=1}^R (-1)^{R-r} \frac{r^{\alpha R}}{\prod_{j=0}^{r-1}(r^\alpha - j^\alpha)\prod_{j=r+1}^{R}(j^\alpha - r^\alpha)} \frac{1}{r^{\alpha R}} = \frac{(-1)^{R-1}}{R!^\alpha} \tag{2.18}$$

and

$$\epsilon_{R+1}(n) = \frac{1}{\tilde{\mathbf{w}}_{R+1}} \sum_{r=1}^R \frac{(-1)^{R-r}}{\prod_{j=0}^{r-1}(r^\alpha - j^\alpha)\prod_{j=r+1}^{R}(j^\alpha - r^\alpha)} \epsilon_r(n) \to 0, \quad \text{as} \quad n \to +\infty. \tag{2.19}$$

We now approximate the new target $\sum_{r=1}^R \mathbf{w}_r\theta^{*,rn}$, by means of $M \in \mathbb{N}^*$ steps of R SA schemes which write

$$\forall r \in [\![1, R]\!], \quad \theta_{p+1}^{rn} = \theta_p^{rn} - \gamma_{p+1} H(\theta_p^{rn}, (U^{rn})^{p+1}), \; p \in [\![0, M-1]\!] \tag{2.20}$$

where $((U^{rn})^p, r = 1, \cdots, R)_{p\in[\![1,M]\!]}$ is an i.i.d sequence with the same law as $(U^{rn}, r = 1, \cdots, R)$, θ_0^{rn}, $r = 1\cdots, R$ are the initial conditions independent of the innovation sequence satisfying $\sup_{n\geq 1}\mathbb{E}|\theta_0^n|^2 < +\infty$ and the sequence $(\gamma_p)_{p\geq 1}$ satisfies (1.2). Now the new *statistical error* of the Richardson-Romberg extrapolation estimator writes

$$\mathcal{E}_S^{R-R}(n, M, R, \gamma, T, H) := \sum_{r=1}^{R} \mathbf{w}_r(\theta^{*,rn} - \theta_M^{rn}).$$

We are looking for an efficient estimator among the family $\left\{\sum_{r=1}^{R} \mathbf{w}_r\theta_M^{rn}, (n, M) \in (\mathbb{N}^*)^2\right\}$. To be more precise, we will minimize the computational cost for a given $L^1(\mathbb{P})$-error $\varepsilon > 0$. We assume that the cost of a single simulation of U^n is proportional to n and is given by $K \times n$, where K is a generic positive constant independent of n. It notably corresponds to the case of discretization schemes of a stochastic process. In the case of the Richardson-Romberg method for SA, at each step $p \in [\![1, M]\!]$ of the procedure, for every $r \in [\![1, R]\!]$, one has to simulate the random vector $(U^n, U^{2n}, \cdots, U^{Rn})$ so that the global computational cost is given by

$$\text{Cost(R-R)} := KM\sum_{r=1}^{R} rn = KMn\frac{R(R+1)}{2}.$$

Hence the problem of interest writes

$$(n(\epsilon), M(\epsilon)) = \arg \min_{\mathbb{E}|\mathcal{E}_{glob}^{R-R}|\leq\varepsilon} \text{Cost(R-R)}.$$

From a practical point of view the constraint: $\mathbb{E}|\mathcal{E}_{glob}^{R-R}| \leq \varepsilon$ is not tractable since one does not have any explicit control on $\mathbb{E}|\mathcal{E}_{glob}^{R-R}|$. Hence one is led to consider some sharp upper bound of this $L^1(\mathbb{P})$-norm, namely

$$\begin{aligned}\mathbb{E}|\mathcal{E}_{glob}^{R-R}| &\leq \left|\sum_{r=1}^{R} \mathbf{w}_r\theta^{*,rn} - \theta^*\right| + \mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r(\theta^{*,rn} - \theta_M^{rn})\right|\right] \\ &\leq \frac{|C_R|}{(R!n^R)^\alpha}\left(1 + |\epsilon_{R+1}(n)|\right) + \mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r(\theta^{*,rn} - \theta_M^{rn})\right|\right]. \end{aligned} \tag{2.21}$$

Note that the bound (2.21) is not tractable since we do not have any closed form expression for the last term appearing in the right-hand side, namely the L^1-norm (or L^2-norm) of the statistical error of the Richardson-Romberg SA estimator. Again we will consider some sharp upper bound. In order to derive an explicit control we assume that the following conditions are in force:

(HUI) $\exists\delta > 0$, such that $\forall\theta \in \mathbb{R}^d$, $\sup_{n\in\mathbb{N}^*} \mathbb{E}[|H(\theta, U^n)|^{2+\delta}] < +\infty$.
(HC1) $\exists C > 0$ such that $\forall n \in \mathbb{N}^*, \forall\theta \in \mathbb{R}^d$, $\mathbb{E}[|H(\theta, U^n)|^2] \leq C(1 + |\theta - \theta^{*,n}|^2)$.
(HC2) $\forall\theta \in \mathbb{R}^d$, $\mathbb{P}(U \notin \mathcal{C}_\theta) = 0$ with $\mathcal{C}_\theta := \{x \in \mathbb{R}^q : x \mapsto H(\theta, x) \text{ is continuous at } x\}$.
(HRG) There exists $a \in (0, 1]$,

$$\sup_{n\in\mathbb{N}^*,(\theta,\theta')\in(\mathbb{R}^d)^2} \frac{\mathbb{E}|H(\theta, U^n) - H(\theta', U^n)|^2}{|\theta - \theta'|^{2a}} < +\infty.$$

(HUA) For each $n \in \mathbb{N}^*$, the map $h^n : \theta \in \mathbb{R}^d \mapsto \mathbb{E}[H(\theta, U^n)]$ is continuously differentiable with Dh^n Lipschitz-continuous uniformly in n and there exists $\underline{\lambda} > 0$ s.t. $\inf_{n\in\mathbb{N}^*,\theta\in\mathbb{R}^d} \lambda_{min}\left((Dh^n(\theta) + Dh^n(\theta)^T)/2\right) > \underline{\lambda}$ where $\lambda_{min}(A)$ denotes the lowest eigenvalue of the matrix A. (*Uniform Attractivity*).
(HS) The step sequence is given by $\gamma_p = \gamma(p)$, $p \geq 1$, where γ is a positive function defined on $[0, +\infty[$ decreasing to zero satisfying one of the following assumptions:

- γ varies regularly with exponent $(-\rho)$, $\rho \in (1/2, 1)$, that is, for any $x > 0$, $\lim_{t\to+\infty} \gamma(tx)/\gamma(t) = x^{-\rho}$.
- for $t \geq 1$, $\gamma(t) = \gamma_0/t$ and γ_0 satisfies $2\underline{\lambda}\gamma_0 > 1$.

Remark 2.1. *Assumption* **(HUA)** *already appears in [Duf96] and [BMP90], see also [FM12] and [FF13] in another context. It allows to control the L^2-norm $\mathbb{E}|\theta_p^{rn} - \theta^{*,rn}|^2$, $r \in [\![1, R]\!]$ with respect to the step $\gamma(p)$ uniformly in n, see section 5, lemma 5.2 . As discussed in [KY03], (Chapter 10, Section 5, p.350, Theorem 5.2) if one considers the projected version of the algorithm* (1.3) *on a bounded convex set D, namely*

$$\theta_{p+1}^n = \Pi_D \left[\theta_p^n - \gamma_{p+1} H(\theta_p^n, (U^n)^{p+1})\right], \; p \in [\![0, M-1]\!],$$

where Π_D denotes the orthogonal projection operator on D (for instance one may set $D = \Pi_{i=1}^d [a_i, b_i]$, $-\infty < a_i < b_i < +\infty$) and $\forall n \geq 1$, $\theta^{*,n} \in int(D)$, as very often happens from a practical point of view, then assumption **(HUA)** *can be localized on D, that is $\inf_{n\in\mathbb{N}^*, \theta\in D} \lambda_{min}\left((Dh^n(\theta) + Dh^n(\theta)^T)/2\right) > \underline{\lambda}$.*

We also want to point out that if assumption **(HUA)** *is satisfied then passing to the limit as $n \to +\infty$ one easily shows that $\lambda_{min}\left((Dh(\theta^*) + Dh(\theta^*)^T)/2\right) \geq \underline{\lambda}$.*

Proposition 2.3. *($L^1(\mathbb{P})$ control of the statistical error) Let $R \in \mathbb{N}^*$. Suppose that for $r \in [\![1, R]\!]$, $U^{rn} \xrightarrow{\mathbb{P}} U^r$ and $\theta_0^n \xrightarrow{\mathbb{P}} \theta_0$, as $n \to +\infty$. Under* **(H-R)**, **(HUI)**, **(HC1)**, **(HC2)**, **(HRG)**, **(HS)** *and* **(HUA)**, *one has for some positive constant $C := C(\gamma, \underline{\lambda})$*

$$\mathbb{E}[|\mathcal{E}_S^{R-R}|] \leq C\mathbb{E}\left[\left|\sum_{r=1}^R \mathbf{w}_r H(\theta^*, U^r)\right|^2\right]^{1/2} \gamma^{1/2}(M)\left(1 + \phi_1^R(n) + \phi_2^R(M)\right)$$

where ϕ_1^R, ϕ_2^R are two positive functions satisfying: $\phi_1^R(n) \to 0$ and $\phi_2^R(M) \to 0$ respectively as $M \to +\infty$, $n \to +\infty$ and ϕ_2^R is non-increasing.

Proof. We define for all $p \geq 1$, $\Delta M_p^{rn} := h^{rn}(\theta_{p-1}^{rn}) - H(\theta_{p-1}^{rn}, (U^{rn})^p) = \mathbb{E}[H(\theta_{p-1}^{rn}, (U^{rn})^p) \big| \mathcal{F}_{p-1}] - H(\theta_{p-1}^{rn}, (U^{rn})^p)$. Recalling that $((U^n, U^{2n}, \cdots, U^{rn}, \cdots, U^{Rn})^p)_{p\in[\![1,M]\!]}$ is a sequence of i.i.d. random variables we have that $(\Delta M_p^{rn})_{p\geq 1}$, $r \in [\![1, R]\!]$, are sequences of martingale increments w.r.t. the natural filtration of the stochastic approximation schemes $\mathcal{F} := (\mathcal{F}_p := \sigma(\theta_0^{rn}, (U^{rn})^1, \cdots, (U^{rn})^p, r = 1, \cdots, R); p \geq 1)$. Using Taylor's formula we get for $p \geq 0$ and $r \in [\![1, R]\!]$

$$\begin{aligned}
\theta_{p+1}^{rn} - \theta^{*,rn} &= \theta_p^{rn} - \theta^{*,rn} - \gamma_{p+1} h^{rn}(\theta_p^{rn}) + \gamma_{p+1}\Delta M_{p+1}^{rn} \\
&= \theta_p^{rn} - \theta^{*,rn} - \gamma_{p+1} Dh(\theta^*)(\theta_p^{rn} - \theta^{*,rn}) \\
&+ \gamma_{p+1}\left(Dh(\theta^*) - \int_0^1 d\lambda Dh^{rn}(\theta^{*,rn} + (1-\lambda)(\theta_p^{rn} - \theta^{*,rn}))\right)(\theta_p^{rn} - \theta^{*,rn}) + \gamma_{p+1}\Delta M_{p+1}^{rn}.
\end{aligned}$$

Hence by a simple induction argument one has for $(r, M) \in [\![1, R]\!] \times \mathbb{N}^*$

$$\theta_M^{rn} - \theta^{*,rn} = \Pi_{1,M}(\theta_0^{rn} - \theta^{*,rn}) + \sum_{k=1}^M \gamma_k \Pi_{k+1,M}\Delta M_k^{rn} + \sum_{k=1}^M \gamma_k \Pi_{k+1,M} R_{k-1}^{rn} \tag{2.22}$$

where $R_k^{rn} = \left(Dh(\theta^*) - \int_0^1 d\lambda Dh^{rn}(\theta^{*,rn} + (1-\lambda)(\theta_k^{rn} - \theta^{*,rn}))\right)(\theta_k^{rn} - \theta^{*,rn})$ and $\Pi_{k,M} := \prod_{j=k}^M (I_d - \gamma_j Dh(\theta^*))$, with the convention that $\Pi_{M+1,M} = I_d$. Multiplying (2.22) on the left by $\mathbf{w}_r$ given by (2.16) and summing w.r.t r lead to

$$\mathcal{E}_S^{R-R} = \Pi_{1,M}\left(\sum_{r=1}^R \mathbf{w}_r(\theta_0^{rn} - \theta^{*,rn})\right) + \sum_{k=1}^M \gamma_k \Pi_{k+1,M}\left(\sum_{r=1}^R \mathbf{w}_r \Delta M_k^{rn}\right) + \sum_{k=1}^M \gamma_k \Pi_{k+1,M}\left(\sum_{r=1}^R \mathbf{w}_r R_{k-1}^{rn}\right) \tag{2.23}$$

Ought to the Minkowski inequality it is sufficient to bound the $L^1(\mathbb{P})$-norm of each term in the above decomposition. First, since $-Dh(\theta^*)$ is a Hurwitz matrix, $\forall \lambda \in [0, \underline{\lambda})$, there exists $C > 0$ such that for any $k \leq n$, $\|\Pi_{k,n}\| \leq C \prod_{j=k}^{n}(1 - \lambda\gamma_j) \leq C \exp(-\lambda \sum_{j=k}^{n} \gamma_j)$. We refer to [Duf96] and [BMP90] for more details. Hence, one has for all $\eta \in (0, \underline{\lambda})$

$$\mathbb{E}[|\Pi_{1,M}(\sum_{r=1}^{R} \mathbf{w}_r(\theta_0^{rn} - \theta^{*,rn}))|] \leq ||\Pi_{1,M}||\mathbb{E}[|\sum_{r=1}^{R} \mathbf{w}_r(\theta_0^{rn} - \theta^{*,rn})|] \leq e^{-(\underline{\lambda}-\eta)\sum_{k=1}^{M}\gamma_k}\mathbb{E}[|\sum_{r=1}^{R} \mathbf{w}_r(\theta_0^{rn} - \theta^{*,rn})|].$$

where $||.||$ stands for the matrix norm on $\mathbb{R}^d \otimes \mathbb{R}^d$. For the second term, recalling that $\sum_{r=1}^{R} \mathbf{w}_r \Delta M_k^{rn}$ is a martingale increment, one has

$$\mathbb{E}\left[\left|\sum_{k=1}^{M} \gamma_k \Pi_{k+1,M}(\sum_{r=1}^{R} \mathbf{w}_r \Delta M_k^{rn})\right|^2\right]^{1/2} \leq \left(\sum_{k=1}^{M} \gamma_k^2 ||\Pi_{1,M}||^2 \mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r \Delta M_k^{rn}\right|^2\right]\right)^{1/2}. \tag{2.24}$$

Similarly for the last term, one has

$$\mathbb{E}\left[\left|\sum_{k=1}^{M} \gamma_k \Pi_{k+1,M}\left(\sum_{r=1}^{R} \mathbf{w}_r R_{k-1}^{rn}\right)\right|\right] \leq \sum_{k=1}^{M} \gamma_k ||\Pi_{1,M}||\mathbb{E}\left|\sum_{r=1}^{R} \mathbf{w}_r R_{k-1}^{rn}\right|. \tag{2.25}$$

We now study the limit of each bound as n and M go to infinity. For the first term, observe that $\sum_{r=1}^{R} \mathbf{w}_r(\theta_0^{rn} - \theta^{*,rn}) \xrightarrow{\mathbb{P}} \sum_{r=1}^{R} \mathbf{w}_r(\theta_0 - \theta^*) = \theta_0 - \theta^*$ as $n \to +\infty$. Moreover, since $\sup_{n\geq 1} \mathbb{E}|\theta_0^n|^2 < +\infty$, by uniform integrability one has $\mathbb{E}|\sum_{r=1}^{R} \mathbf{w}_r(\theta_0^{rn} - \theta^{*,rn})| \to \mathbb{E}|\theta_0 - \theta^*|$ as $n \to +\infty$. If $\gamma(p) = \gamma_0/p$ we select η such that $2(\underline{\lambda} - \eta)\gamma_0 > 1$ otherwise we set $\eta < \underline{\lambda}$. Hence we get

$$\mathbb{E}[|\Pi_{1,M}(\sum_{r=1}^{R} \mathbf{w}_r(\theta_0^{rn} - \theta^{*,rn}))|] \leq \gamma^{1/2}(M)\mathbb{E}\left[|\theta_0 - \theta^*|\right]\left(1 + \phi_1^R(n)\right)$$

with $\phi_1^R(n) \to 0$ as $n \to +\infty$. Let us now study the second term. Define for $k \geq 1$, $\Delta N_k^{rn} = h^{rn}(\theta^*) - H(\theta^*, (U^{rn})^k)$ then by the Cauchy-Schwarz inequality and **(HRG)** one has

$$\begin{aligned}
\left|\mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r \Delta M_k^{rn}\right|^2\right] - \mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r \Delta N_k^{rn}\right|^2\right]\right| &\leq C_R\left(\sum_{r=1}^{R} ||\mathbf{w}_r||\mathbb{E}\left[|\Delta M_k^{rn} - \Delta N_k^{rn}|^2\right]\right)^{1/2} \\
&\quad \times \left(\mathbb{E}[|H(\theta_{k-1}^{rn}, (U^{rn})^k)|^2]^{1/2} + \mathbb{E}[|H(\theta^*, (U^{rn})^k)|^2]^{1/2}\right) \\
&\leq C_R \max_{1\leq r\leq R} \mathbb{E}[|\theta_{k-1}^{rn} - \theta^*|^{2a}]^{1/2} \\
&\leq C_R(\gamma_k^{a/2} + n^{-a\alpha})
\end{aligned}$$

where we used lemma 5.2 and $\max_{1\leq r\leq R} |\theta^{*,rn} - \theta^*| \leq Cn^{-\alpha}$ for the last inequality. Now observe that $\mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r \Delta N_k^{rn}\right|^2\right] = \mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r(h^{rn}(\theta^*) - H(\theta^*, U^{rn}))\right|^2\right]$ so that using **(HC2)** and $U^{rn} \xrightarrow{\mathbb{P}} U^r$ as $n \to +\infty$, one has $\sum_{r=1}^{R} \mathbf{w}_r(h^{rn}(\theta^*) - H(\theta^*, U^{rn})) \xrightarrow{\mathbb{P}} \sum_{r=1}^{R} \mathbf{w}_r H(\theta^*, U^r)$ as $n \to +\infty$. From **(HUI)** we

deduce the L^2-uniform integrability of the family $\left\{\sum_{r=1}^R \mathbf{w}_r(h^{rn}(\theta^*) - H(\theta^*, U^{rn})), n \geq 1\right\}$ which yields

$$\mathbb{E}\left[\left|\sum_{r=1}^R \mathbf{w}_r \Delta N_k^{rn}\right|^2\right] \longrightarrow \mathbb{E}\left[\left|\sum_{r=1}^R \mathbf{w}_r H(\theta^*, U^r)\right|^2\right], \quad n \to +\infty.$$

Plugging the above estimates into (2.24), we derive the following bound

$$\mathbb{E}\left[\left|\sum_{k=1}^M \gamma_k \Pi_{k+1,M}(\sum_{r=1}^R \mathbf{w}_r \Delta M_k^{rn})\right|^2\right]^{1/2} \leq \mathbb{E}\left[\left|\sum_{r=1}^R \mathbf{w}_r H(\theta^*, U^r)\right|^2\right]^{1/2} \left(\sum_{k=1}^M \gamma_k^2 ||\Pi_{1,M}||^2\right)^{1/2} (1 + \phi_1^R(n)) \tag{2.26}$$
$$+ C_R \left(\sum_{k=1}^M \gamma_k^2 \gamma_k^{a/2} ||\Pi_{1,M}||^2\right)^{1/2},$$

with $\phi_1^R(n) \to 0$ as $n \to +\infty$. Using lemma 5.1, we successively derive that $\left(\sum_{k=1}^M \gamma_k^2 ||\Pi_{1,M}||^2\right)^{1/2} \leq C\gamma^{1/2}(M)$ for some positive constant $C(\gamma, \underline{\lambda})$ and $\left(\sum_{k=1}^M \gamma_k^2 \gamma_k^{a/2} ||\Pi_{1,M}||^2\right)^{1/2} = o(\gamma^{1/2}(M)) = \gamma^{1/2}(M)\phi_2^R(M)$ as $M \to +\infty$. We now focus on the last term. Let us first observe that using **(H-R)** and since Dh^{rn} is Lipschitz (uniformly in n) one has

$$\begin{aligned}|R_k^{rn}| &= \left|\left(Dh(\theta^*) - Dh^{rn}(\theta^*) + \int_0^1 d\lambda \left(Dh^{rn}(\theta^*) - Dh^{rn}(\theta^{*,rn} + (1-\lambda)(\theta_k^{rn} - \theta^{*,rn}))\right)\right)(\theta_k^{rn} - \theta^{*,rn})\right| \\ &\leq C\left(\max_{1\leq r\leq R} ||Dh(\theta^*) - Dh^{rn}(\theta^*)|| + |\theta_k^{rn} - \theta^{*,rn}|\right)|\theta_k^{rn} - \theta^{*,rn}|\end{aligned}$$

so that plugging this estimate in (2.25) and using lemma 5.2 lead to

$$\mathbb{E}\left[\left|\sum_{k=1}^M \gamma_k \Pi_{k+1,M}\left(\sum_{r=1}^R \mathbf{w}_r R_{k-1}^{rn}\right)\right|\right] \leq C\left(\sum_{k=1}^M (\gamma_k^{3/2} \max_{1\leq r\leq R} ||Dh(\theta^*) - Dh^{rn}(\theta^*)|| + \gamma_k^2)||\Pi_{1,M}||\right).$$

Finally lemma 5.1 and since $\max_{1\leq r\leq R} ||Dh(\theta^*) - Dh^{rn}(\theta^*)|| \to 0$ as $n \to +\infty$ also imply

$$\max_{1\leq r\leq R} ||Dh(\theta^*) - Dh^{rn}(\theta^*)|| \sum_{k=1}^M \gamma_k^{3/2} ||\Pi_{1,M}|| \leq C\gamma^{1/2}(M)\phi_1^R(n)$$

and

$$\sum_{k=1}^M \gamma_k^2 ||\Pi_{1,M}|| = o(\gamma^{1/2}(M)) = \gamma^{1/2}(M)\phi_2^R(M).$$

□

From the previous computations we are naturally led to consider the following suboptimal computational cost optimization problem

$$(n(\epsilon), M(\epsilon)) = \arg \min_{\mu_R n^{-\alpha R}(1+|\epsilon_{R+1}(n)|)+\nu_R\gamma^{1/2}(M)(1+\phi_1^R(n)+\phi_2^R(M))\leq\varepsilon} \text{Cost(R-R)} \tag{2.27}$$

where $\mu_R = \frac{|C_R|}{R!^\alpha}$ and $\nu_R = C\mathbb{E}\left[\left|\sum_{r=1}^R \mathbf{w}_r H(\theta^*, U^r)\right|^2\right]^{1/2}$.

Proposition 2.4. *(Computational cost optimization) Let $R \in \mathbb{N}^*$. Suppose that the assumptions of Proposition 2.3 are satisfied. Suppose that the step sequence γ is given by: $\gamma(p) = \gamma_0/p^\beta$, $\gamma_0 > 0$, $p > 0$, $\beta \in (1/2, 1]$. The multi-step Richardson-Romberg SA estimator of order R satisfies*

$$\inf_{\mu_R n^{-\alpha R}(1+|\epsilon_{R+1}(n)|)+\nu_R\gamma^{1/2}(M)(1+\phi_1^R(n)+\phi_2^R(M))\leq\varepsilon} \text{Cost(R-R)} \sim K\frac{R(R+1)}{2}\gamma_0^{\frac{1}{\beta}}\nu_R^{\frac{2}{\beta}}\mu_R^{\frac{1}{\alpha R}}\frac{1}{\varepsilon^{\frac{2}{\beta}+\frac{1}{\alpha R}}}\left(1+\frac{2\alpha R}{\beta}\right)^{\frac{1}{\alpha R}}\left(1+\frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}}$$

as $\varepsilon \to 0$. Eventually this asymptotically optimal bound may be achieved with parameters satisfying:

$$n(\varepsilon) \sim \left(\frac{2\alpha R}{\beta}+1\right)^{\frac{1}{\alpha R}}\mu_R^{\frac{1}{\alpha R}}\varepsilon^{-\frac{1}{\alpha R}} \quad \text{and} \quad M(\varepsilon) \sim \gamma_0^{\frac{1}{\beta}}\nu_R^{\frac{2}{\beta}}\left(1+\frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}}\varepsilon^{-\frac{2}{\beta}} \quad \text{as } \varepsilon \to 0. \tag{2.28}$$

Proof. Let us note that the cost minimization problem (2.27) is lower-bounded by the more tractable problem

$$\inf_{\mu_R n^{-\alpha R}+\nu_R\gamma^{1/2}(M)\leq\varepsilon} \text{Cost(R-R)} = \inf_{\mu_R n^{-\alpha R}<\varepsilon} K\gamma^{-1}\left(\frac{(\epsilon-\mu_R n^{-\alpha R})}{\nu_R^2}\right)n\frac{R(R+1)}{2} \tag{2.29}$$

with $M = \gamma^{-1}\left(\frac{(\varepsilon-\mu_R n^{-\alpha R})^2}{\nu_R^2}\right) = \gamma_0^{1/\beta}\nu_R^{2/\beta}(\varepsilon-\mu_R n^{-\alpha R})^{-2/\beta}$. This optimization problem can be solved explicitly, more precisely the optimal parameters are given by

$$n(\varepsilon) = \left(\frac{2\alpha R}{\beta}+1\right)^{\frac{1}{\alpha R}}\mu_R^{\frac{1}{\alpha R}}\varepsilon^{-\frac{1}{\alpha R}}, \quad M(\varepsilon) = \gamma_0^{\frac{1}{\beta}}\nu_R^{\frac{2}{\beta}}\left(1+\frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}}\varepsilon^{-\frac{2}{\beta}}.$$

The "liminf" side of the result clearly follows by plugging this solution into (2.29). Now set

$$n(\varepsilon) = \left(\frac{2\alpha R}{\beta}+1\right)^{\frac{1}{\alpha R}}\mu_R^{\frac{1}{\alpha R}}\varepsilon^{-\frac{1}{\alpha R}}, \quad M(\varepsilon) = \gamma^{-1}\left(\frac{(\varepsilon-\mu_R(1+|\varepsilon_{R+1}(n(\varepsilon)))n^{-\alpha R}(\varepsilon))^2}{\nu_R^2\left(1+\phi_1^R(n(\varepsilon))+\phi_2^R(\gamma_0^{\frac{1}{\beta}}\nu_R^{\frac{2}{\beta}}\left(1+\frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}}\varepsilon^{-\frac{2}{\beta}})\right)^2}\right).$$

Since ϕ_2^R is non-increasing, the couple $(n(\varepsilon), M(\varepsilon))$ satisfies the constraint $\mu_R n^{-\alpha R}(1+|\epsilon_{R+1}(n)|)+\nu_R\gamma^{1/2}(M)(1+\phi_1^R(n)+\phi_2^R(M)) \leq \varepsilon$ so that the cost minimization problem (2.27) is upper-bounded by

$$K\frac{R(R+1)}{2}\mu_R^{\frac{1}{\alpha R}}\varepsilon^{-\frac{1}{\alpha R}}\left(1+\frac{2\alpha R}{\beta}\right)^{\frac{1}{\alpha R}}\gamma_0^{\frac{1}{\beta}}\nu_R^{\frac{2}{\beta}}\left(1+\frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}}\varepsilon^{-\frac{2}{\beta}}\left(1-(1+|\varepsilon_{R+1}(n(\varepsilon)))\frac{\beta}{2\alpha R+\beta}\right)^{\frac{2}{\beta}}$$
$$\times\left(1+\phi_1^R(n(\varepsilon))+\phi_R^2(\gamma_0^{\frac{1}{\beta}}\nu_R^{\frac{2}{\beta}}\left(1+\frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}}\varepsilon^{-\frac{2}{\beta}})\right)^{\frac{2}{\beta}}$$

and the result follows by letting ε goes to zero.

□

Remark 2.2. ***(Choice of the step sequence)*** *According to Proposition 2.4, it is optimal to set $\beta = 1$ to achieve a minimal asymptotic complexity. In this case a constraint appear on γ_0: $2\underline{\lambda}\gamma_0 > 1$. Let us note that for $\beta = 1$ a simple computation shows that the constant C appearing in ν_R is equal to $\gamma_0/(2\underline{\lambda}\gamma_0-1)^{1/2}$ which reaches its minimum (as a function of γ_0) at $\gamma_0 = 1/\underline{\lambda}$. However the main drawback with this choice is that the constant $\underline{\lambda}$ is not known to the experimenter so that one is led to make a blind choice in practical implementation.*

Remark 2.3. ***(Control of the variance)*** *Let us note that when one decides to implement the Richardson-Romberg extrapolation SA scheme with an innovation satisfying $U^r = U$ a.s. $r = 1, \cdots, R$ then one has $H(\theta^*, U^r) = H(\theta^*, U)$ a.s. for every $r \in [\![1, R]\!]$ so that using (2.14) yields*

$$\mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r H(\theta^*, U^r)\right|^2\right] = \mathbb{E}\left[\left|(\sum_{r=1}^{R} \mathbf{w}_r) H(\theta^*, U)\right|^2\right] = \mathbb{E}\left[|H(\theta^*, U)|^2\right].$$

Hence we clearly see that this choice leads to a control in the L^1-norm of the statistical error of the multi-step Richardson-Romberg SA estimator. On the opposite considering mutually independent innovations U^r lead to an explosion of the previous control with respect to R. Indeed one has

$$\begin{aligned}
\mathbb{E}\left[\left|\sum_{r=1}^{R} \mathbf{w}_r H(\theta^*, U^r)\right|^2\right] &= \left(\sum_{r=1}^{R} \frac{r^{2\alpha R}}{\prod_{j=0}^{r-1}(r^\alpha - j^\alpha)^2 \prod_{j=r+1}^{R}(j^\alpha - r^\alpha)^2}\right) \mathbb{E}\left[|H(\theta^*, U)|^2\right] \\
&\geq \left(\frac{R^R}{R!}\right)^{2\alpha} \mathbb{E}\left[|H(\theta^*, U)|^2\right] \\
&\sim \left(\frac{e^R}{\sqrt{2\pi}\sqrt{R}}\right)^{2\alpha} \mathbb{E}\left[|H(\theta^*, U)|^2\right] \quad \text{as } R \to +\infty,
\end{aligned}$$

where we used (2.16) for the first equality.

For instance when one is concerned with the discretization of a Brownian diffusion, the first aforementioned case consists in implementing the Richardson-Romberg method with R Euler schemes devised with the same Brownian motion *W namely $W^r = W$, $r = 1, \cdots, R$ whereas the second case consists in implementing the method with mutually independent Brownian motions W^r. The optimality of this choice is discussed in [Pag07].*

2.3. Comparison with the crude stochastic approximation estimator

Under the assumptions of Proposition 2.3 with $R = 1$, the global error for the crude stochastic approximation estimator satisfies

$$\begin{aligned}
\mathbb{E}\left[|\mathcal{E}_{glob}(M, \gamma, H)|\right] = \mathbb{E}\left[|\theta^* - \theta^{*,n} + \theta^{*,n} - \theta_M^n|\right] &\leq \frac{|C_1|}{n^\alpha}(1 + |\varepsilon_1(n)|) + \mathbb{E}\left[|\theta^{*,n} - \theta_M^n|\right] \\
&\leq \frac{|C_1|}{n^\alpha}(1 + |\epsilon_1(n)|) + C\mathbb{E}\left[|H(\theta^*, U)|^2\right]^{\frac{1}{2}} \gamma^{\frac{1}{2}}(M)(1 + \phi_1(n) + \phi_2(M)),
\end{aligned}$$

with a computational cost given by Cost(C-S) $:= KMn$. Hence a similar result as in Proposition 2.4 holds.

Proposition 2.5. *Assume that the assumptions of Proposition 2.3 with $R = 1$ hold. Suppose that the step sequence γ is given by: $\gamma(p) = \gamma_0/p^\beta$, $\gamma_0 > 0$, $p > 0$, $\beta \in (1/2, 1]$. The crude stochastic approximation estimator satisfies*

$$\inf_{|C_1|n^{-\alpha}(1+|\epsilon_1(n)|)+\nu_1\gamma^{1/2}(M)(1+\phi_1(n)+\phi_2(M))\leq\varepsilon} \text{Cost(C-S)} \sim K\gamma_0^{\frac{1}{\beta}}\nu_1^{\frac{2}{\beta}}|C_1|^{\frac{1}{\alpha}} \frac{1}{\varepsilon^{\frac{2}{\beta}+\frac{1}{\alpha}}} \left(1 + \frac{2\alpha}{\beta}\right)^{\frac{1}{\alpha}} \left(1 + \frac{\beta}{2\alpha}\right)^{\frac{2}{\beta}}$$

as $\varepsilon \to 0$ with $\nu_1 = C\mathbb{E}\left[|H(\theta^*, U)|^2\right]^{\frac{1}{2}}$. Eventually this asymptotically optimal bound may be achieved with parameters satisfying:

$$n(\varepsilon) \sim \left(\frac{2\alpha}{\beta} + 1\right)^{\frac{1}{\alpha}} |C_1|^{\frac{1}{\alpha}} \varepsilon^{-\frac{1}{\alpha}} \quad \text{and} \quad M(\varepsilon) \sim \gamma_0^{\frac{1}{\beta}} \nu_1^{\frac{2}{\beta}} \left(1 + \frac{\beta}{2\alpha}\right)^{\frac{2}{\beta}} \varepsilon^{-\frac{2}{\beta}} \quad \text{as } \varepsilon \to 0. \tag{2.30}$$

3. Application: Estimation of the quantile of a component of a SDE

In this section, we show how the previous results can be applied to the estimation of the quantile of a stochastic process solution to a stochastic differential equation. Also, when the exact value of a constant is not important we may repeat the same symbol for constants that may change from one line to next.

3.1. Notations and Hypotheses.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions and $(Z_t)_{t\geq 0}$ be a q-dimensional $(\mathcal{F}_t)_{t\geq 0}$ symmetric α-stable process, for $\alpha \in (0,2]$, that is a càdlàg process with independent and stationary increments with the scaling property $Z_{ct} \overset{(d)}{=} c^{1/\alpha} Z_t$. Note that the case $\alpha = 2$ corresponds to the standard Brownian motion. It is also the only case where Z is a continuous process. When $\alpha < 2$, the Stable process is discontinuous and its Lévy-Khintchine exponent writes for all $p \in \mathbb{R}^q$,

$$\mathbb{E}\left(e^{i\langle p, Z_t\rangle}\right) = \exp\left(-t\int_{S^{q-1}} |\langle p, \vartheta\rangle|^\alpha \mu(d\vartheta)\right).$$

We refer to the measure μ as the *spectral measure* of Z. It is related to the Lévy measure of the process Z as follows. Denote ν the Lévy measure of Z, ν factorizes in $\nu(dz) = C_\alpha \frac{d|z|}{|z|^{1+\alpha}}\mu(\bar{z})$, where $z = (|z|, \bar{z}) \in \mathbb{R}_+ \times S^{d-1}$ stands for the polar coordinates. For the exact value of C_α, we refer to Sato [Sat05]. Let us consider a d-dimensional process $(X_t)_{t\geq 0} = (X_t^1, \ldots, X_t^d)_{t\geq 0}$ with dynamics:

$$X_t = x + \int_0^t b(X_{s-})ds + \int_0^t \sigma(X_{s-})dZ_s, \tag{3.31}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^q$. We fix the time horizon $T = 1$. Let us denote by $\mathbb{P}_x$ (resp. $\mathbb{P}_{t,x}$, $t \in (0,1]$) the conditional probability given $\{X_0 = x\}$ (resp. $\{X_t = x\}$). For a given level $\ell \in (0,1)$, we are interested in the computation of the quantile at level ℓ of the random variable X_1^d defined as:

$$\theta^* = \inf\{\theta \in \mathbb{R} : \mathbb{P}_x(X_1^d \leq \theta) \geq \ell\}.$$

Since $\lim_{\theta\to+\infty} \mathbb{P}_x(X_1^d \leq \theta) = 1$, we have $\{\theta \in \mathbb{R} : \mathbb{P}_x(X_1^d \leq \theta) \geq \ell\} \neq \emptyset$. Moreover, we have $\lim_{\theta\to-\infty} \mathbb{P}_x(X_1^d \leq \theta) = 0$, which implies that $\{\theta \in \mathbb{R} : \mathbb{P}_x(X_1^d \leq \theta) \geq \ell\}$ is bounded from below so that θ^* always exists. Assuming that the distribution of X_1^d has no atoms, the quantile at level ℓ is the lowest solution of the equation:

$$\mathbb{P}_x(X_1^d \leq \theta) = \ell.$$

If the distribution function is (strictly) increasing, which is notably the case if the process X solution of (3.31) admits a positive density $p(1, x, .)$, the solution to the above equation is unique, otherwise, there may be more than one solution. Now since the law of X_1^d is not known explicitly, the quantile θ^* cannot be computed and one has to approximate the dynamics by a discretization scheme that can be simulated. Let us note that the estimation of the quantile of a component of a Brownian diffusion process has already been investigated in [TZ04]. For a given time step $\Delta = \frac{1}{n}$, $n \in \mathbb{N}^*$, setting for all $i \in \mathbb{N}$, $t_i = i\Delta$, we consider the standard Euler scheme defined as follows:

$$X_t^n = x + \int_0^t b(X_{\phi(s)}^n)ds + \int_0^t \sigma(X_{\phi(s)}^n)dZ_s \quad \phi(s) = \sup\{t_i : \ t_i \leq s\}. \tag{3.32}$$

Then one approximates θ^* by $\theta^{*,n}$ the quantile at level ℓ of $X_1^{n,d}$. We denote by $[\mathbf{A}]$ the following set of assumptions. Fix an integer $m \in \mathbb{N}$, for $\alpha \in (0,2)$ we assume that $m \geq d+4$, which will hereafter refer to the regularity of the coefficients.

[A-1] $b \in \mathcal{C}^m(\mathbb{R}^d, \mathbb{R}^d)$ and $\sigma \in \mathcal{C}^m(\mathbb{R}^d, \mathbb{R}^d \otimes \mathbb{R}^q)$ with bounded derivatives. Also, when $\alpha \leq 1$, we put $b = 0$.
[A-2] For all $x, \xi \in \mathbb{R}^q$, there exists $0 < c \leq C$ such that:

$$c|\xi|^2 \leq \langle \xi, \sigma(x)\xi \rangle \leq C|\xi|^2. \tag{3.33}$$

[A-3] When $\alpha < 2$, the spectral measure μ has $\mathcal{C}^m(S^{q-1})$ surface density and satisfies: for all $\xi \in \mathbb{R}^q$, there exists $C_1, C_2 > 0$ such that:

$$C_1|\xi|^\alpha \leq \int_{S^{q-1}} |\langle \xi, \vartheta \rangle|^\alpha \mu(d\vartheta) \leq C_2|\xi|^\alpha. \tag{3.34}$$

Proposition 3.1. *Assume that $\alpha \in (0, 2]$ and that **[A]** is in force. For every $t > 0$, the solutions X_t, X_t^n, of the SDE* (3.31) *and* (3.32) *have a strictly positive densities with respect to the Lebesgue measure. Consequently, the quantile is uniquely defined. Moreover, those densities are in $\mathcal{C}^m(\mathbb{R}^d, \mathbb{R}^d)$ if $\alpha > 1$, and in $\mathcal{C}^{m-1}(\mathbb{R}^d, \mathbb{R}^d)$ when $\alpha \leq 1$.*

We refer to the work of Kolokoltsov [Kol00] for the proof in the Stable case, who also derived Aronson's estimates with time singularity depending on the index α. In the Brownian case, i.e. $\alpha = 2$, if the drift b is a measurable bounded function and the diffusion coefficient σ is η-Hölder continuous, $\eta > 0$, and satisfies **[A-2]** then the aforementioned densities exists, are positive and satisfy Gaussian Aronson's estimates (see e.g. [Fri64] and [LM10] for the density of the Euler scheme).

Proposition 3.2. *For $\alpha \in (0, 2)$ assume that **[A]** for $m = 2$. For $\alpha = 2$, assume the drift b and the diffusion coefficient σ are Lipschitz-continuous bounded functions and that σ satisfies **[A-2]**.Then one has*

$$\theta^{*,n} \to \theta^*, \ n \to +\infty.$$

Proof. Let $n \in \mathbb{N}^*$ and denote by F, F^n the distribution function of X_1^d and $X_1^{n,d}$ respectively. Since b and σ are Lipschitz we know that $(X_1^{n,d})_{n\geq 1}$ converges in distribution to X_1^d. Moreover, the function F is continuous so that $(F^n)_{n\geq 1}$ converges uniformly to F. Hence, we conclude that $F(\theta^{*,n}) \to \ell$, $n \to +\infty$. Now remark that from Proposition 3.1 since X_1^d has a strictly positive density the function F is one-to-one which in turn implies that F^{-1} exists and is continuous so that $\theta^{*,n} \to F^{-1}(\ell) = \theta^*$. □

From Proposition 3.1 (existence of a positive density for $X_1^{n,d}$) the quantile $\theta^{*,n}$ at level ℓ of the random variable $X_1^{n,d}$ is the unique solution of the equation

$$\mathbb{P}_x\left(X_1^{n,d} \leq \theta\right) = \ell.$$

In this section, we are interested in giving an expansion for the error $\theta^* - \theta^{*,n}$ in powers of n^{-1}, using Theorem 2.1. Actually, we will prove that **[A]** implies **[H-k]**, for a desired $k > 0$. As we can see, Theorem 2.1 requires an expansion of $h^n - h$ and its derivatives up to order $k > 0$ in order to have an expansion of $\theta^* - \theta^{*,n}$ at the same order. Regularity of the function h may be obtained mainly by two means: either the function H is smooth w.r.t. the variable θ (with polynomial growth w.r.t θ and x) or the laws of X_T and X_T^n are smooth. Concerning the expansion of the difference $\partial_\theta^k h - \partial_\theta^k h^n$ it may also be obtained by two means: in the regular setting i.e. when the function $x \mapsto \partial_\theta^k H(\theta, x)$ and the coefficients b and σ are regular (say b, σ, $\partial_\theta^k H(\theta, .)$ are $\mathcal{C}_b^{R+5}$) one may use standard tools such as the one developed in Talay-Tubaro [TT90] (in the Brownian case); or in the (Hypo-)elliptic setting, the laws of X_T and X_T^n are smooth. Here, we are in the latter case. Indeed, the estimation of the quantile of a diffusion can be seen as an inverse problem, by setting $H(\theta, x) = 1 - \frac{1}{1-\ell}\mathbf{1}_{\{x^d \geq \theta\}}$. We thus see that regularity of H fails. However, for $\theta \in \mathbb{R}$, we have:

$$h(\theta) - h^n(\theta) = \frac{1}{1-\ell}\left(\mathbb{P}^x(X_1^d \leq \theta) - \mathbb{P}^x(X_1^{n,d} \leq \theta)\right).$$

Let $p(T,x,\theta)$ be the density of the diffusion, and $p_n(T,x,\theta)$ the density of the Euler scheme at time T. The derivative w.r.t. θ of the previous equality is:

$$\forall k \geq 1, \forall(\theta,x) \in \mathbb{R}\times\mathbb{R}^d,\ \frac{d^k}{d\theta^k}h(\theta) - \frac{d^k}{d\theta^k}h^n(\theta) = \frac{1}{1-\ell}\left(\frac{\partial^{k-1}}{\partial\theta^{k-1}}p^{X_1^d}(1,x,\theta) - \frac{\partial^{k-1}}{\partial\theta^{k-1}}p_n^{X_1^{n,d}}(1,x,\theta)\right),$$

where we denote by $p^{X_1^d}(1,x,\theta)$ and $p_n^{X_1^{n,d}}(1,x,\theta)$ the marginal densities of X_1^d and $X_1^{n,d}$. Consequently, we observe that in order to apply Theorem 2.1, we have to give an expansion of the marginal densities and their derivatives, up to an order $k>1$. Actually, we will show that the expansion holds for $p(1,x,\theta)-p_n(1,x,\theta)$ and its derivatives, the expansion for the marginals will follow from an integration over the $d-1$ first components.

3.2. **Expansion for the densities.**

Using a continuity technique known as the Parametrix expansion, Konavov and Mammen [KM02], in the Brownian case, and Konakov and Menozzi [KM11], in the stable case, successfully derive an expansion for the density of the solution of (3.31) to an arbitrary order, with explicit terms. The purpose of this section is to extend these results to the derivatives of the densities.

The Parametrix expansion consists in representing the density of the solution of (3.31) as a series involving the density of a *frozen* equation and the generators associated with (3.31) and the *frozen* density. We take a few lines here to describe this technique.

We define the following process as the frozen process. Recall $T=1$ is a fixed deterministic time. For a given terminal point $y\in\mathbb{R}^d$, the frozen equation at point y is defined as:

$$\tilde{X}_t = x + b(y)t + \sigma(y)Z_t. \tag{3.35}$$

Thanks to the uniform ellipticity of σ, the process (3.35) has a density with respect to the Lebesgue measure, given by:

$$\tilde{p}_\alpha^y(t,x,y) = \begin{cases} \dfrac{\det(\Sigma(y))^{-1/2}}{(2\pi t)^{d/2}}\exp\left(-\dfrac{1}{2t}(y-x-b(y)t)^T\Sigma(y)^{-1}(y-x-b(y)t)\right), & \text{if } \alpha=2 \\ \dfrac{1}{(2\pi)^d}\displaystyle\int_{\mathbb{R}^d} dp e^{-i\langle p,y-x-b(y)t\rangle}\exp\left(-t\int_{S^{d-1}}|\langle p,\sigma(y)\vartheta\rangle|^\alpha\mu(d\vartheta)\right), & \text{if } \alpha\in(0,2). \end{cases}$$

We will often drop the superscript y with the convention $\tilde{p}_\alpha(t,x,y)=\tilde{p}_\alpha^y(t,x,y)$, when no ambiguity is possible. The distance between $p(t,x,y)$ and $\tilde{p}_\alpha(t,x,y)$ will then be quantified by the difference of the generators of (3.31) and (3.35). We put $\Sigma(z)=\sigma(z)\sigma(z)^T$. The generator of the SDE (3.31):

$$Lf(t,x,y) = \begin{cases} \dfrac{1}{2}\mathrm{tr}\left(\Sigma(x)\partial_x^2 f(t,x,y)\right) + \langle b(x),\partial_x f(t,x,y)\rangle, & \text{if } \alpha=2\ , \\ \langle b(x),\partial_x f(t,x,y)\rangle + \displaystyle\int_{\mathbb{R}^d} f(t,x+\sigma(x)z,y) - f(t,x,y) - \frac{\langle\nabla_x f(t,x,y),\sigma(x)z\rangle}{1+|z|^2}\nu(dz), & \text{if } \alpha\in(0,2). \end{cases}$$

The integro-differential part can be written :

$$\int_{\mathbb{R}^d} f(t,x+\sigma(x)z,y) - f(t,x,y) - \frac{\langle\nabla_x f(t,x,y),\sigma(x)z\rangle}{1+|z|^2}\nu(dz) = \int_{S^{d-1}}|\langle\partial_x,\sigma(x)\vartheta\rangle|^\alpha f(t,x,y)\mu(d\vartheta),$$

where the notation $\int_{S^{d-1}}|\langle\partial_x,\sigma(x)\vartheta\rangle|^\alpha f(t,x,y)\mu(d\vartheta)$ stands for a fractional derivative and has to be understood as a Fourier multiplier:

$$\int_{S^{d-1}}|\langle\partial_x,\sigma(x)\vartheta\rangle|^\alpha f(t,x,y)\mu(d\vartheta) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle p,x\rangle}\int_{S^{d-1}}|\langle p,\sigma(x)\vartheta\rangle|^\alpha\mu(d\vartheta)\mathcal{F}_x(f)(t,p,y)dp,$$

where $\mathcal{F}_x(f)(t,p,y)$ denotes the Fourier transform with respect to the first space argument of the function $f(t,x,y)$. Let us define the generator of the frozen process (3.35):

$$\tilde{L}^* f(t,x,y) = \begin{cases} \frac{1}{2}\mathrm{tr}\left(\Sigma(y)\partial_x^2 f(t,x,y)\right) + \langle b(y), \partial_x f(t,x,y)\rangle, & \text{if } \alpha = 2, \\ \langle b(y), \partial_x f(t,x,y)\rangle - \int_{S^{d-1}} |\langle \partial_x, \sigma(y)\vartheta\rangle|^\alpha f(t,x,y)\mu(d\vartheta), & \text{if } \alpha \in (0,2). \end{cases}$$

Theorem 3.1. *Under the assumptions [**A**], the solution of* (3.31) *exists and has density with respect to the Lebesgue measure. Let $p(t,x,y)$ denote the density of* (3.31)*. It admits the following representation:*

$$p(t,x,y) = \sum_{k=0}^{\infty} \tilde{p}_\alpha \otimes H^{(k)}(t,x,y),$$

where we denoted $H(t,x,y) = (L - \tilde{L}^*)\tilde{p}_\alpha(t,x,y)$, and $\otimes$ is the space-time convolution:

$$f \otimes g\,(t,x,y) = \int_0^t \int_{\mathbb{R}} f\,(u,x,z)\, g\,(t-u,z,y)\, dz du,$$

and $H^{(k)}(t,x,y) = H^{(k-1)} \otimes H(t,x,y)$, where $H^{(0)}(t,x,y) = Id$.

This result has been investigated in the literature, let us mention Friedman [Fri64] for the Brownian case and Kolokoltsov [Kol00] for the stable case. The proof relies on a precise study of the frozen density and its derivatives (fractional derivatives in the stable case), and show that in the time space convolution, the time singularities induced by the derivation can be compensated to get a convergent series.

Similarly, one gets an equivalent result for the density of the Euler scheme. We introduce the "frozen Markov chains" $(\tilde{X}^n_{t_k})_{k\in[\![0,n]\!]}$:

$$\tilde{X}^n_{t_k} = x,\ \tilde{X}^n_{t_{k+1}} = \tilde{X}^n_{t_k} + b(y)\Delta + \sigma(y)(Z_{t_{k+1}} - Z_{t_k}).$$

We denote the discrete generators:

$$L_n f(t_k - t_j, x, y) = \Delta^{-1}\left(\int p_n(\Delta,x,z) f(t_k - t_{j+1}, z, y)dz - f(t_k - t_{j+1}, x, y)\right), \tag{3.36}$$

$$\tilde{L}^*_n f(t_k - t_j, x, y) = \Delta^{-1}\left(\int \tilde{p}^y(\Delta,x,z) f(t_k - t_{j+1}, z, y)dz - f(t_k - t_{j+1}, x, y)\right). \tag{3.37}$$

We then obtain a representation of the density of the Euler scheme using the frozen density and the discrete generators.

Theorem 3.2. *The density $p_n(t_k,x,y)$ of the Euler scheme admits the following representation:*

$$p_n(t_k - t_j, x, y) = \sum_{k=0}^{k-j} \tilde{p}_\alpha \otimes_n H_n^{(k,n)}(t_k - t_j, x, y),$$

where we denoted $H_n(t_k,x,y) = (L_n - \tilde{L}_n)\tilde{p}(t_k,x,y)$, and $\otimes_n$ is the discretized space-time convolution:

$$f \otimes_n g\,(t_k,x,y) = \frac{1}{n}\sum_{i=0}^{n-1} \int_{\mathbb{R}} f\,(t_i,x,z)\, g\,(t_k - t_i, z, y)\, dz,$$

and $H_n^{(r,n)}(t_k,x,y) = H_n^{(r-1,n)} \otimes_n H_n(t_k,x,y)$, where $H_n^{(0,n)}(t_k,x,y) = Id$.

Once again, these results have been investigated in the literature and we state them here without proof. The reader may consult [KM02, KM11] and the references therein.

Roughly speaking, we see that the differences between the two expansions of Theorems 3.1 and 3.2 come from the convolution and the kernel. Thus, in order to get an expansion for $p-p_n$, we introduce for all $k \in [\![0, n-1]\!]$:

$$\begin{aligned} p^d(t_k,x,y) &= \sum_{r=0}^{+\infty} \tilde{p}_\alpha \otimes_n H^{(r,n)}(t_k,x,y), \text{ where } H(t,x,y) = (L-\tilde{L}^*)\tilde{p}_\alpha(t,x,y), \\ H^{(r,n)}(t_k,x,y) &= H^{(r-1,n)} \otimes_n H(t_k,x,y), \text{ where } H^{(0,n)}(t_k,x,y) = Id. \end{aligned}$$

Formally, p^d is the series of Theorem 3.1, with the time integrals discretized. We then look for an expansion for the two differences $p-p_n = p-p^d+p^d-p_n$. To that end, we define $\tilde{L}_* f(t,x,y) = \tilde{L}_x f(t,x,y)$, where :

$$\tilde{L}_\xi f(t,x,y) = \begin{cases} \frac{1}{2}\mathrm{tr}\left(\Sigma(\xi)\partial_x^2 f(t,x,y)\right) + \langle b(\xi), \partial_x f(t,x,y)\rangle, & \text{if } \alpha = 2, \\ \langle b(\xi), \partial_x f(t,x,y)\rangle - \int_{S^{d-1}} |\langle \partial_x, \sigma(\xi)\vartheta\rangle|^\alpha f(t,x,y)\mu(d\vartheta), & \text{if } \alpha \in (0,2). \end{cases}$$

Note that both generators $\tilde{L}^*$ and $\tilde{L}_*$ depends on the freezing parameter y. This induces extra caution below, as we will be led to differentiate with respect to the freezing parameter.

Extending the results of Theorem 1.1 in Konakov and Mammen in [KM02], for the Brownian case, and Theorem 21 in Konakov and Menozzi in [KM11], for the Stable case, we have the following result.

Theorem 3.3. *Assume that [**A**] holds. Let $M \in \mathbb{N}^*$ be such that when $\alpha = 2$, $0 < M \le m/2$, and when $\alpha < 2$, $0 < M \le m-(d+4)$. Let $\gamma \in \mathbb{N}^d$, with $|\gamma| \le M$. Then, for all $x, y \in \mathbb{R}^d$, we have:*

$$\begin{aligned} \partial_y^\gamma p(1,x,y) - \partial_y^\gamma p_n(1,x,y) &= \sum_{k=1}^{M-1-|\gamma|} \frac{1}{(k+1)!n^k}\partial_y^\gamma \left(p \otimes_n \left(L-\tilde{L}^*\right)^{k+1} p^d\right)(1,x,y) \\ &\quad - \frac{1}{(k+1)!n^k}\partial_y^\gamma \left(p^d \otimes_n \left(\tilde{L}_*-\tilde{L}^*\right)^{k+1} p_n\right)(1,x,y) + \frac{\partial_y^\gamma R(x,y)}{n^{M-|\gamma|}}. \end{aligned} \tag{3.38}$$

*Also, there is a constant $C > 0$ depending on the set of assumptions [**A**], T, γ, and M such that the following bound holds for each term and the remainders:*

$$\begin{aligned} \sum_{k=1}^{M-|\gamma|-1} \left|\partial_y^\gamma \left(p \otimes_n \left(L-\tilde{L}^*\right)^{k+1} p^d\right)(1,x,y)\right| + \left|\partial_y^\gamma \left(p^d \otimes_n \left(\tilde{L}_*-\tilde{L}^*\right)^{k+1} p_n\right)(1,x,y)\right| \\ + |\partial_y^\gamma R(x,y)| \le C\bar{p}_K^\alpha(t,x,y), \end{aligned} \tag{3.39}$$

where for a given $K > 0$, we denoted $\bar{p}_K^\alpha(t,x,y)$ the following quantity:

$$\bar{p}_K^\alpha(t,x,y) = \begin{cases} t^{-d/2}\exp\left(-K\frac{|y-x|^2}{t}\right), & \text{if } \alpha = 2, \\ \frac{t^{-d/\alpha}}{\left[K \vee \frac{|y-x|}{t^{\frac{1}{\alpha}}}\right]^{d+\alpha}}, & \text{if } \alpha \in (0,2). \end{cases}$$

For $\gamma = 0$, expansion (3.38) is given in [KM02] in the Brownian case, and in [KM11] in the stable case. To get an expansion for $\partial_y^\gamma(p-p_n)(1,x,y)$, we take the derivative along y in each term in that expansion, and prove that each one is bounded by an α stable density.

Formally, $\bar{p}^{\alpha}_{K}(t,x,y)$ is a stable density (up to some normalizing constant depending on $K>0$). Observe that $\bar{p}^{\alpha}$ satisfies a semi-group property in the following sense:

Proposition 3.3. *For all $\tau\in(0,t)$, for all $x,y\in\mathbb{R}^d$ for all $K_1,K_2>0$, there exists $K,C>0$ depending on the set of assumptions [**A**] and the terminal time T, such that:*

$$\int_{\mathbb{R}^d}\bar{p}^{\alpha}_{K_1}(\tau,x,z)\bar{p}^{\alpha}_{K_2}(t-\tau,z,y)dz\leq C\bar{p}^{\alpha}_{K}(t,x,y). \tag{3.40}$$

Proof. Indeed, for all $\alpha\in(0,2]$, we have that for $t>0$, for all $x,y\in\mathbb{R}^d$, there exists $c,C,K>0$ such that:

$$c\bar{p}^{\alpha}_{K}(t,x,y)\leq\tilde{p}^{y}_{\alpha}(t,x,y)\leq C\bar{p}^{\alpha}_{K}(t,x,y). \tag{3.41}$$

For the gaussian case, we refer to the seminal paper [Fri64] or Sheu [She91] for a stochastic control based approach. For the stable case $\alpha<2$, the reader may consult and Kolokolstov [Kol00]. Thus, one easily gets:

$$\int_{\mathbb{R}^d}\bar{p}^{\alpha}_{K_1}(\tau,x,z)\bar{p}^{\alpha}_{K_2}(t-\tau,z,y)dz\leq C\int_{\mathbb{R}^d}\tilde{p}^{y}_{\alpha}(\tau,x,z)\tilde{p}^{y}_{\alpha}(t-\tau,z,y)dz\leq C\tilde{p}^{y}_{\alpha}(t,x,y)\leq C\bar{p}^{\alpha}_{K}(t,x,y).$$

□

Using the previous density, we are able to bound the various terms appearing above.

Lemma 3.1. *For all multi index $\gamma,\eta\in\mathbb{N}^d$ such that $|\gamma|+|\eta|\leq m$ if $\alpha>1$, and $|\gamma|+|\eta|\leq m-1$ if $\alpha\leq 1$, for all $x,y\in\mathbb{R}^d$, for all $t\in[0,T]$, for all $k\in[\![0,n-1]\!]$, there exists $C=C([\mathbf{A}],T,\gamma,\eta)>0$ such that the following bounds holds:*

$$|\partial^{\gamma}_{x}\partial^{\eta}_{y}p^{d}(t_k,x,y)|+|\partial^{\gamma}_{x}\partial^{\eta}_{y}p_{n}(t_k,x,y)|\leq Ct_k^{-\frac{|\gamma|+|\eta|}{\alpha}}\bar{p}^{\alpha}_{K}(t_k,x,y), \tag{3.42}$$

$$|\partial^{\gamma}_{x}\partial^{\eta}_{y}p(t,x,y)|\leq Ct^{-\frac{|\gamma|+|\eta|}{\alpha}}\bar{p}^{\alpha}_{K}(t,x,y), \tag{3.43}$$

Moreover, for all $\xi\in\mathbb{R}^d$,

$$|\partial^{\gamma}_{x}p^{d}(t_k,x,x+\xi)|+|\partial^{\gamma}_{x}p_{n}(t_k,x,x+\xi)|\leq C\bar{p}^{\alpha}_{K}(t_k,x,x+\xi). \tag{3.44}$$

Eventually, when $\alpha<2$, denoting $\Phi(t_k,x,y)=\sum_{r=1}^{\infty}H^{(r,n)}(t_k,x,y)$, we have:

$$\left|\partial^{\gamma}_{x}\partial^{\eta}_{y}\Phi(t_k,x,y)\right| \leq Ct^{-\frac{|\gamma|+|\eta|}{\alpha}}\bar{p}^{\alpha}_{K}(t_k,x,y)\left(1+\frac{1\wedge|x-y|}{t_k}\right), \tag{3.45}$$

$$|\partial^{\gamma}_{x}\Phi(t_k,x,x+\xi)| \leq C\bar{p}^{\alpha}_{K}(t_k,x,x+\xi)\left(1+\frac{1\wedge|\xi|}{t_k}\right). \tag{3.46}$$

Remark 3.1. *We point out that in equations (3.44) and (3.46), despite the presence of derivations, there are no singularities induced by them, as the argument appears in both the forward and the backward arguments. This will be a key point in the proof of Theorem 3.3.*

Proof. For the Brownian case, all the above estimates are proved in [KM02]. We thus focus on the stable case. In Konakov Menozzi [KM11], the bound (3.42) and (3.43) are given. To get the bound (3.44), we prove (3.45) and (3.46), using the following estimates proved in [KM11]:

$$\left|\partial^{\gamma}_{x}\partial^{\eta}_{y}H(t,x,y)\right| \leq C_1t^{-\frac{|\gamma|+|\eta|}{\alpha}}\bar{p}^{\alpha}_{K}(t,x,y)\left(1+\frac{1\wedge|x-y|}{t}\right), \tag{3.47}$$

$$|\partial^{\gamma}_{x}H(t,x,x+\xi)| \leq C_1\bar{p}^{\alpha}_{K}(t,x,x+\xi)\left(1+\frac{1\wedge|\xi|}{t}\right). \tag{3.48}$$

We then derive (3.44) for the derivative of the densities using the expansion:

$$p^d(1,x,y) = \tilde{p}_\alpha(t,x,y) + \frac{1}{n}\sum_{i=0}^{n-1}\int_{\mathbb{R}^d} \tilde{p}_\alpha\left(t_i,x,z\right)\Phi\left(1-t_i,z,y\right)dz. \tag{3.49}$$

We begin with (3.46). Observe that due to the presence of the derivation parameter in both the forward and the backward arguments, the derivatives does not yield any additional singularities. From (3.48), we prove by induction the following:

$$\left|\partial_x^\gamma H^{(r,n)}(t_k,x,x+\xi)\right| \leq C_r t_k^{(r-1)\omega}\bar{p}_K^\alpha(t_k,x,x+\xi)\left(1+\frac{1\wedge|\xi|}{t_k}\right), \tag{3.50}$$

where $\omega = \frac{1}{\alpha}\wedge\alpha$, and the sequence of constants $(C_r)_{r\geq 0}$ is defined recursively by:

$$C_{r+1} = C_\gamma C_r C \max\left(\frac{1}{r\omega}, B\big((r-1)\omega+1,\omega\big)\right),\ C_1>0,$$

where C_1 is the constant appearing in bounds (3.48) and (3.50), and C is a positive constant independent of r,γ,x,ξ. For $r=1$, the bound is exactly (3.48). Suppose that it holds for $r\geq 1$. We have using the induction hypothesis, equation (3.48) and Leibnitz's formula:

$$\begin{aligned}
\left|\partial_x^\gamma H^{(r+1,n)}(t_k,x,x+\xi)\right| &\leq \sum_{\eta=0}^{\gamma} C_\gamma^\eta \frac{1}{n}\sum_{i=0}^{k-1}\int_{\mathbb{R}^d}\left|\partial_x^\eta H^{(r,n)}(t_i,x,z+x)\right|\left|\partial_x^{\gamma-\eta}H(t_k-t_i,z+x,x+\xi)\right|dz\\
&\leq C_\gamma C_r C\frac{1}{n}\sum_{i=0}^{k-1}\int_{\mathbb{R}^d} t_i^{(r-1)\omega}\bar{p}_K^\alpha(t_i,x,x+z)\left(1+\frac{1\wedge|z|}{t_i}\right)\\
&\quad\times\ \bar{p}_K^\alpha(t_k-t_i,x+z,x+\xi)\left(1+\frac{1\wedge|\xi-z|}{t_k-t_i}\right)dz.
\end{aligned}\tag{3.51}$$

We decompose, the integral:

$$\int_{\mathbb{R}^d}\bar{p}_K^\alpha(t_i,x,x+z)\left(1+\frac{1\wedge|z|}{t_i}\right)\bar{p}_K^\alpha(t_k-t_i,x+z,x+\xi)\left(1+\frac{1\wedge|\xi-z|}{t_k-t_i}\right)dz = I_1+I_2+I_3+I_4 \tag{3.52}$$

where:

$$\begin{aligned}
I_1 &= \int_{\mathbb{R}^d}\bar{p}_K^\alpha(t_i,x,x+z)\bar{p}_K^\alpha(t_k-t_i,x+z,x+\xi)dz,\\
I_2 &= \int_{\mathbb{R}^d}\bar{p}_K^\alpha(t_i,x,x+z)\bar{p}_K^\alpha(t_k-t_i,x+z,x+\xi)\frac{1\wedge|\xi-z|}{t_k-t_i}dz\\
I_3 &= \int_{\mathbb{R}^d}\bar{p}_K^\alpha(t_i,x,x+z)\frac{1\wedge|z|}{t_i}\bar{p}_K^\alpha(t_k-t_i,x+z,x+\xi)dz,\\
I_4 &= \int_{\mathbb{R}^d}\bar{p}_K^\alpha(t_i,x,x+z)\frac{1\wedge|z|}{t_i}\bar{p}_K^\alpha(t_k-t_i,x+z,x+\xi)\frac{1\wedge|\xi-z|}{t_k-t_i}dz.
\end{aligned}$$

The first one I_1 is bounded by $C\bar{p}_K^\alpha(t_k,x,x+\xi)$ thanks to the semi-group property (Proposition 3.3). By symmetry, I_2 and I_3 are treated the same way. We focus on I_2. In the rest, we denote by the symbol $\asymp$ the relation:

$$f(x)\asymp g(x) \Leftrightarrow \exists C>1,\ \forall x\in\mathbb{R}^d:\ C^{-1}g(x)\leq f(x)\leq Cg(x).$$

We argue differently, according to the ratio $|\xi|/t_k^{1/\alpha}$.

- Suppose first that $|\xi| \leq t_k^{1/\alpha}$. Then, the diagonal estimate holds: $\bar{p}_K^\alpha(t_k, x, x+\xi) \asymp t_k^{-d/\alpha}$. On the one hand, if $i \geq k/2$, then $t_i \asymp t_k$. Since the diagonal estimate is a global bound, one has: $\bar{p}_K^\alpha(t_i, x, x+z) \leq Ct_i^{-d/\alpha} \asymp Ct_k^{-d/\alpha} \asymp C\bar{p}_K^\alpha(t_k, x, x+\xi)$. On the other hand, when $i \leq k/2$, then $t_k - t_i \asymp t_k$, and we have $\frac{1}{t_k - t_i}\bar{p}_K^\alpha(t_k - t_i, x+z, x+\xi) \leq C\frac{1}{t_k - t_i}(t_k - t_i)^{-d/\alpha} \asymp C\frac{1}{t_k}\bar{p}_K^\alpha(t_k, x, x+\xi)$.
- Suppose now that $|\xi| \geq t_k^{1/\alpha}$. Then, the diagonal estimate holds: $\bar{p}_K^\alpha(t_k, x, x+\xi) \asymp \frac{t_k}{|\xi|^{d+\alpha}}$. Now, since $|\xi| \leq |z| + |\xi - z|$, we have either $1/2|\xi| \leq |z|$, or $1/2|\xi| \leq |\xi - z|$. In the first case the off-diagonal estimate holds for the first density: $\bar{p}_K^\alpha(t_i, x, x+z) \asymp \frac{t_i}{|z|^{d+\alpha}} \leq C\frac{t_k}{|\xi|^{d+\alpha}} \asymp C\bar{p}_K^\alpha(t_k, x, x+\xi)$. In the second case, the second density is off-diagonal and we can write: $\frac{1}{t_k - t_i}\bar{p}_K^\alpha(t_k - t_i, x+z, x+\xi) \leq C\frac{1}{t_k - t_i}\frac{t_k - t_i}{|\xi - z|^{d+\alpha}} \leq \frac{1}{t_k}\frac{t_k}{|\xi|^{d+\alpha}} \asymp C\frac{1}{t_k}\bar{p}_K^\alpha(t_k, x, x+\xi)$.

Therefore, we always have the alternative:

$$\bar{p}_K^\alpha(t_i, x, x+z) \leq C\bar{p}_K^\alpha(t_k, x, x+\xi), \text{ or } \frac{1}{t_k - t_i}\bar{p}_K^\alpha(t_k - t_i, x+z, x+\xi) \leq C\frac{1}{t_k}\bar{p}_K^\alpha(t_k, x, x+\xi). \tag{3.53}$$

Combining this alternative with the smoothing effect of the Parametrix kernel H reflected in the bound (see Section 3 of Kolokoltsov [Kol00]):

$$\forall \tau \in (0, T),\ \forall y \in \mathbb{R}^d,\ \int_{\mathbb{R}^d} \frac{1 \wedge |y - z|}{\tau}\bar{p}_K^\alpha(\tau, z, y)dz \leq \tau^{(\alpha \wedge 1) - 1}, \tag{3.54}$$

gives that the second and third terms are bounded by:

$$I_2 + I_3 \leq C\bar{p}_K^\alpha(t_k, x, x+\xi)\left(t_i^{(\alpha\wedge 1)-1} + (t_k - t_i)^{(\alpha \wedge 1)-1} + \frac{1 \wedge |\xi|}{t_k}\right).$$

We now turn to the last term in (3.52), that writes:

$$I_4 = \int_{\mathbb{R}^d} \frac{1 \wedge |z|}{t_i}\bar{p}_K^\alpha(t_i, x, x+z)\frac{1 \wedge |\xi - z|}{t_k - t_i}\bar{p}_K^\alpha(t_k - t_i, x+z, x+\xi)dz.$$

When $\bar{p}_K^\alpha(t_k, x, x+\xi)$ is in the diagonal regime, that is, when $|\xi| \leq Ct_k^{\frac{1}{\alpha}}$, we have:

$$\frac{1 \wedge |z|}{t_i}\bar{p}_K^\alpha(t_i, x, x+z) \leq Ct_i^{-d/\alpha}t_i^{\frac{1}{\alpha}-1} \text{ and } \frac{1 \wedge |\xi - z|}{t_k - t_i}\bar{p}_K^\alpha(t_k - t_i, x+z, x+\xi) \leq C(t_k - t_i)^{-d/\alpha}(t_k - t_i)^{\frac{1}{\alpha}-1}.$$

We prove the first inequality, the second one is obtained with the same arguments. Let us assume first that $|z| \leq Ct_i^{1/\alpha}$. In that case the diagonal estimate holds for $\bar{p}_K^\alpha(t_i, x, x+z)$, thus:

$$\frac{1 \wedge |z|}{t_i}\bar{p}_K^\alpha(t_i, x, x+z) \leq \frac{1 \wedge |z|}{t_i}t_i^{-d/\alpha} \leq \frac{|z|}{t_i}t_i^{-d/\alpha} \leq Ct_i^{1/\alpha - 1} \times t_i^{-d/\alpha}.$$

On the other hand, when the off-diagonal estimate holds for $\bar{p}_K^\alpha(t_i, x, x+z)$, that is when $|z| > Ct_i^{1/\alpha}$, we have:

$$\frac{1 \wedge |z|}{t_i}\bar{p}_K^\alpha(t_i, x, x+z) \leq \frac{1 \wedge |z|}{t_i}\frac{t_i}{|z|^{d+\alpha}} \leq C\frac{1}{|z|^{d+\alpha-1}} \leq Ct_i^{-\frac{1}{\alpha}(d+\alpha-1)} = Ct_i^{-\frac{d}{\alpha}}t_i^{\frac{1}{\alpha}-1}.$$

Thus, in both cases, we obtained the announced bound.

Now, if $i \leq k/2$, $t_k \asymp t_k - t_i$, one has:

$$\frac{1 \wedge |\xi - z|}{t_k - t_i}\bar{p}_K^\alpha(t_k - t_i, x+z, x+\xi) \leq C\bar{p}_K^\alpha(t_k, x, x+\xi)(t_k - t_i)^{\frac{1}{\alpha}-1}.$$

Then, using (3.54), $I_4 \leq \bar{p}_K^\alpha(t_k, x, x+\xi)(t_k - t_i)^{\frac{1}{\alpha}-1} t_i^{(\alpha\wedge 1)-1}$. Similarly, when $i > k/2$, we use that $t_k \asymp t_i$, to get

$$\frac{1\wedge|z|}{t_i}\bar{p}_K^\alpha(t_i, x, x+z) \leq C\bar{p}_K^\alpha(t_k, x, x+\xi) t_i^{\frac{1}{\alpha}-1}.$$

Consequently, when $|\xi| \leq t_k^{\frac{1}{\alpha}}$, I_4 is bounded in the following way:

$$I_4 \leq C\bar{p}_K^\alpha(t_k, x, x+\xi)\Big((t_k - t_i)^{\frac{1}{\alpha}-1} t_i^{(\alpha\wedge 1)-1} + t_i^{\frac{1}{\alpha}-1}(t_k - t_i)^{(\alpha\wedge 1)-1}\Big).$$

Assume now that $|\xi| > Ct_k^{\frac{1}{\alpha}}$. In that case using similar arguments one may prove that we have either:

$$\frac{1\wedge|\xi - z|}{t_k - t_i}\bar{p}_K^\alpha(t_k - t_i, x+z, x+\xi) \leq C\frac{1\wedge|\xi|}{t_k}\bar{p}_K^\alpha(t_k, x, x+\xi),$$

or:

$$\frac{1\wedge|z|}{t_i}\bar{p}_K^\alpha(t_i, x, x+z) \leq C\frac{1\wedge|\xi|}{t_k}\bar{p}_K^\alpha(t_k, x, x+\xi).$$

Thus, using (3.54), I_4 is now bounded as follows:

$$I_4 \leq C\frac{1\wedge|\xi|}{t_k}\bar{p}_K^\alpha(t_k, x, x+\xi)\Big(t_i^{(\alpha\wedge 1)-1} + (t_k - t_i)^{(\alpha\wedge 1)-1}\Big).$$

Plugging this estimate in (3.52) in turn implies:

$$\begin{aligned}\int_{\mathbb{R}^d} \bar{p}_K^\alpha(t_i, x, x+z)\left(1 + \frac{1\wedge|z|}{t_i}\right)\bar{p}_K^\alpha(t_k - t_i, x+z, x+\xi)\left(1 + \frac{1\wedge|y-z|}{t_k - t_i}\right)dz \\ \leq C\bar{p}_K^\alpha(t_k, x, x+\xi)\Bigg(\Big((t_k - t_i)^{\frac{1}{\alpha}-1} t_i^{(\alpha\wedge 1)-1} + t_i^{\frac{1}{\alpha}-1}(t_k - t_i)^{(\alpha\wedge 1)-1}\Big) \\ + \frac{1\wedge|\xi|}{t_k}\Big(t_i^{(\alpha\wedge 1)-1} + (t_k - t_i)^{(\alpha\wedge 1)-1}\Big)\Bigg). \qquad (3.55)\end{aligned}$$

Plugging bound (3.55) in (3.51) yields:

$$\begin{aligned}\left|\partial_x^\gamma H^{(r+1,n)}(t_k, x, x+\xi)\right| \leq\ & C_\gamma C_r C\bar{p}_K^\alpha(t_k, x, x+\xi)\frac{1}{n}\sum_{i=0}^{k-1} t_i^{(r-1)\omega}\Big((t_k - t_i)^{\frac{1}{\alpha}-1} t_i^{(\alpha\wedge 1)-1} + t_i^{\frac{1}{\alpha}-1}(t_k - t_i)^{(\alpha\wedge 1)-1}\Big) \\ & + C_\gamma C_r C\frac{1\wedge|\xi|}{t_k}\bar{p}_K^\alpha(t_k, x, x+\xi)\frac{1}{n}\sum_{i=0}^{k-1} t_i^{(r-1)\omega}\Big(t_i^{(\alpha\wedge 1)-1} + (t_k - t_i)^{(\alpha\wedge 1)-1}\Big). \qquad (3.56)\end{aligned}$$

Now, assume first that $\alpha \geq 1$. Recalling that $\omega = \frac{1}{\alpha}\wedge\alpha = \frac{1}{\alpha}$, the above bound becomes:

$$\begin{aligned}\left|\partial_x^\gamma H^{(r+1,n)}(t_k,x,x+\xi)\right| &\leq C_\gamma C_r C \bar{p}_K^\alpha(t_k,x,x+\xi)\frac{1}{n}\sum_{i=0}^{k-1} t_i^{(r-1)\frac{1}{\alpha}}\left((t_k-t_i)^{\frac{1}{\alpha}-1}+t_i^{\frac{1}{\alpha}-1}\right)\\ &+ C_\gamma C_r C\frac{1\wedge|\xi|}{t_k}\bar{p}_K^\alpha(t_k,x,x+\xi)\frac{1}{n}\sum_{i=0}^{k-1} t_i^{(r-1)\frac{1}{\alpha}}\\ &\leq C_{r+1} t_k^{\frac{n}{\alpha}}\bar{p}_K^\alpha(t_k,x,x+\xi)\left(1+\frac{1\wedge|\xi|}{t_k}\right),\end{aligned}$$

where $C_{r+1} = C_\gamma C C_r \max\left(B\left((r-1)\frac{1}{\alpha}+1,\frac{1}{\alpha}\right),\frac{\alpha}{r}\right)$ and we used that $t_k^{(r-1)\frac{1}{\alpha}+1} \leq t_k^{\frac{r}{\alpha}}$, since $\alpha \geq 1$ and $t_k \leq 1$.

On the other hand, when $\alpha \leq 1$, $\omega = \alpha$ one similarly proves that:

$$\left|\partial_x^\gamma H^{(r+1,n)}(t_k,x,x+\xi)\right| \leq C_{r+1} t_k^{r\alpha}\bar{p}_K^\alpha(t_k,x,x+\xi)\left(1+\frac{1\wedge|\xi|}{t_k}\right),$$

with $C_{r+1} = C_\gamma C_r C \max\left(\frac{1}{r\alpha}, B\big((r-1)\alpha+1,\alpha\big)\right)$. This constant is coherent with the previous one, setting $C_{r+1} = C_\gamma C_r C \max\left(\frac{1}{r\omega}, B\big((r-1)\omega+1,\omega\big)\right)$. This concludes the proof of bound (3.50). Observe that by definition of Euler's Beta function, $(C_r)_{r\geq 0}$ produces a convergent series. To get the bound (3.46), we sum bounds (3.50). In order to get the bound (3.44), we now plug the bound (3.46) in equation (3.49), and from similar arguments, one derives (3.44).

To prove (3.45), we show by induction the following bound:

$$\left|\partial_x^\gamma\partial_y^\eta H^{(r,n)}(t_k,x,y)\right| \leq C_r t_k^{(r-1)\omega-\frac{|\gamma|+|\eta|}{\alpha}} C_1\bar{p}_K^\alpha(t_k,x,y)\left(1+\frac{1\wedge|x-y|}{t_k}\right). \tag{3.57}$$

For $r=1$, this bound is exactly (3.47). To get the estimate for $r+1$, we proceed as above.

$$\left|\partial_x^\gamma\partial_y^\eta H^{(r+1,n)}(t_k,x,y)\right| = \left|\partial_x^\gamma\partial_y^\eta\frac{1}{n}\sum_{i=0}^{k-1}\int_{\mathbb{R}^d} H^{(r,n)}(t_i,x,z)H(t_k-t_i,z,y)dz\right| \leq I+II,$$

where

$$I = \left|\partial_x^\gamma\partial_y^\eta\frac{1}{n}\sum_{i\leq k/2}\int_{\mathbb{R}^d} H^{(r,n)}(t_i,x,z)H(t_k-t_i,z,y)dz\right|,\ II = \left|\partial_x^\gamma\partial_y^\eta\frac{1}{n}\sum_{i\geq k/2}\int_{\mathbb{R}^d} H^{(r,n)}(t_i,x,z)H(t_k-t_i,z,y)dz\right|.$$

In I, the time parameter t_i is small, thus, the singularities induced by the derivation of $H^{(r,n)}(t_i,x,z)$ are the worst. In order to get rid of them, we make use of a change of variable to get:

$$I = \left|\partial_x^\gamma\frac{1}{n}\sum_{i\leq k/2}\int_{\mathbb{R}^d} H^{(r,n)}(t_i,x,z)\partial_y^\eta H(t_k-t_i,z,y)dz\right| = \left|\partial_x^\gamma\frac{1}{n}\sum_{i\leq k/2}\int_{\mathbb{R}^d} H^{(r,n)}(t_i,x,z+x)\partial_y^\eta H(t_k-t_i,z+x,y)dz\right|.$$

Now, from equations (3.50), (3.47) and Leibnitz's formula we derive:

$$\begin{aligned}
I &\leq \sum_{\beta=0}^{\gamma} C_\gamma^\beta \frac{1}{n} \sum_{i\leq k/2} \int_{\mathbb{R}^d} \left|\partial_x^\beta H^{(r,n)}(t_i,x,z+x)\right| \left|\partial_x^{\gamma-\beta}\partial_y^\eta H(t_k-t_i,z+x,y)\right| dz \\
&\leq C_\gamma C_n \frac{1}{n} \sum_{i\leq k/2} \int_{\mathbb{R}^d} t_i^{(r-1)\omega} \left(1+\frac{1\wedge|z|}{t_i}\right) \bar{p}_K^\alpha(t_i,x,x+z) \\
&\quad \times (t_k-t_i)^{-\frac{|\gamma|+|\eta|}{\alpha}} \left(1+\frac{1\wedge|y-x-z|}{t_k-t_i}\right) \bar{p}_K^\alpha(t_k-t_i,x+z,y)dz. \\
&\leq C_{n+1} t_k^{r\omega-\frac{|\gamma|+|\eta|}{\alpha}} \bar{p}_K^\alpha(t_k,x,y).
\end{aligned}$$

where we used that $t_k \asymp t_k - t_i$ for $i \leq k/2$ for the last inequality. Note that once again, the series $\sum_{r\geq 1} C_r$ converges. For II, we proceed with similar arguments. In this case, we use the change variables $w = z + y$ instead.

□

Proof of Theorem 3.3. The coefficients are the sum of two terms. We only focus on the first term, the second term can be treated similarly. From the definition of $\otimes_n$, we have:

$$\partial_y^\gamma p \otimes_n \left(L-\tilde{L}^*\right)^{k+1} p^d(1,x,y) = \partial_y^\gamma \frac{1}{n} \sum_{i=0}^{n-1} \int_{\mathbb{R}^d} p(t_i,x,z)\big(L-\tilde{L}^*\big)^{k+1} p^d(1-t_i,z,y)dz.$$

To deal with the singularities coming from the derivatives we split the sum over i in two parts:

$$\begin{aligned}
\partial_y^\gamma p \otimes_n \left(L-\tilde{L}^*\right)^{k+1} p^d(1,x,y) &= \partial_y^\gamma \frac{1}{n} \sum_{i<n/2} \int_{\mathbb{R}^d} p(t_i,x,z)\big(L-\tilde{L}^*\big)^{k+1} p^d(1-t_i,z,y)dz \\
&\quad + \partial_y^\gamma \frac{1}{n} \sum_{i\geq n/2} \int_{\mathbb{R}^d} p(t_i,x,z)\big(L-\tilde{L}^*\big)^{k+1} p^d(1-t_i,z,y)dz \\
&= S_1 + S_2.
\end{aligned}$$

For S_1, the time parameter is small, thus the singularities brought by the derivation in y and the generators are negligible. Indeed, exchanging the derivation and the integral:

$$S_1 = \frac{1}{n} \sum_{i<n/2} \int_{\mathbb{R}^d} p(t_i,x,z)\partial_y^\gamma \left(\left(L-\tilde{L}^*\right)^{k+1} p^d(1-t_i,z,y)\right) dz.$$

From bound (3.42) in Lemma 3.1, we derive:

$$\left|\partial_y^\gamma \left(\left(L-\tilde{L}^*\right)^{k+1} p^d(1-t_i,z,y)\right)\right| \leq C\,(1-t_i)^{-k-1-\frac{\gamma}{\alpha}}\, \bar{p}_K^\alpha\,(1-t_i,z,y)\,.$$

The right hand side of the previous equation is bounded uniformly in z, thus, from Lebesgue theorem, we can derive under the integral. Now, since $p(t_i,x,z) \leq C\bar{p}_K^\alpha(t_i,x,z)$, this sum yields by a semi-group property:

$$\begin{aligned}
\left|\partial_y^\gamma \frac{1}{n} \sum_{i<n/2} \int_{\mathbb{R}^d} p(t_i,x,z)\big(L-\tilde{L}^*\big)^{k+1} p^d(1-t_i,z,y)dz\right| &\leq C\frac{1}{n} \sum_{i<n/2} \int_{\mathbb{R}^d} \bar{p}_K^\alpha(t_i,x,z)\,(1-t_i)^{-\frac{\gamma}{\alpha}-k-1}\,\bar{p}_K^\alpha\,(1-t_i,z,y)\,. \\
&= C\bar{p}_K^\alpha(1,x,y).
\end{aligned}$$

We now turn to the second sum. When $i \geq n/2$, by an integration by parts it follows

$$\begin{aligned} S_2 &= \partial_y^\gamma \frac{1}{n} \sum_{i \geq n/2} \int_{\mathbb{R}^d} p(t_i, x, z) \big(L - \tilde{L}^*\big)^{k+1} p^d(1 - t_i, z, y) dz \\ &= \partial_y^\gamma \frac{1}{n} \sum_{i \geq n/2} \int_{\mathbb{R}^d} \Big(\big(L - \tilde{L}^*\big)^{k+1}\Big)^T p(t_i, x, z) p^d(1 - t_i, z, y) dz. \end{aligned}$$

where $\Big(\big(L - \tilde{L}^*\big)^{k+1}\Big)^T$ stands for the adjoint of $\big(L - \tilde{L}^*\big)^{k+1}$, which is well defined thanks to the smoothness of the coefficients b and σ. The differential operator $\Big(\big(L - \tilde{L}^*\big)^{k+1}\Big)^T$ is still a differential operator of order $k+1$, thus, applied to p yields singularities of order $k+1$, which are negligible, since $i \geq n/2$. However, the derivative ∂_y^γ will affect $p^d(1 - t_i, z, y)$, thus, giving additional singularities so that we make use of the change of variable: $z = y - u$ to derive

$$S_2 = \frac{1}{n} \sum_{i \geq n/2} \sum_{\eta=0}^{\gamma} C_\gamma^\eta \int_{\mathbb{R}^d} \partial_y^{\gamma - \eta} \left[\Big(\big(L - \tilde{L}^*\big)^{k+1}\Big)^T p\left(t_i, x, y - u\right) \right] \partial_y^\eta p^d \left(1 - t_i, y - u, y\right) du.$$

Now, bound (3.44) of Lemma 3.1 gives: $\forall \eta \in \mathbb{N}^*$, $\exists C > 0$ s.t.: $\left|\partial_y^\eta p^d \left(1 - t_i, y - u, y\right)\right| \leq C \bar{p}_K^\alpha \left(1 - t_i, y - u, y\right)$. On the other hand, since $i \geq n/2$, the singularities of the derivatives on the first density are negligible. We thus get from a semi group property:

$$|S_2| \leq \frac{C}{n} \sum_{i \geq n/2} \sum_{\eta=0}^{\gamma} C_\gamma^\eta 2^{\frac{|\gamma| + |\eta|}{\alpha} + k + 1} \int_{\mathbb{R}^d} \bar{p}_K^\alpha \left(t_i, x, y - u\right) \bar{p}_K^\alpha \left(1 - t_i, y - u, y\right) du \leq C_\gamma \bar{p}_K^\alpha \left(1, x, y\right).$$

In order to prove that the expansion (3.38) makes sense, it remains to prove the bound on the remainder. Since the expansion (3.38) is made of two contributions $p - p^d$ and $p^d - p_n$, the remainder $R(x, y)$ also splits in two terms $R(x, y) = R_1(1, x, y) + R_2(1, x, y)$, each being a remainder respectively of the expansion of $p - p^d$ and $p^d - p_n$, with:

$$\begin{aligned} R_1(1, x, y) &= \sum_{r \geq 0} (Q_M \otimes_n H^{(r,n)})(1, x, y), \\ Q_M(t_i, x, y) &= \frac{1}{M!} \sum_{i=0}^{k-1} \int_{i/n}^{(i+1)/n} \left[n\left(u - i/n\right)\right] \int_0^1 (1 - \delta)^{M-1} \int \frac{\partial^M}{\partial s^M} \left[p(s, x, z) H(t_i - s, z, y)\right]_{s = t_i + \delta(u - t_i)} dz d\delta du, \end{aligned}$$

and

$$\begin{aligned} R_2(t, x, y) &= \frac{1}{(M+1)!} \int_0^1 (1 - \tau)^M \left[p^d \otimes_n \left(\tilde{L}_* - \tilde{L}^*\right)^{M+1} \tilde{p}_\tau^\Delta \right](t, x, y) \\ \tilde{p}_\tau^\Delta(t, x, y) &= \sum_{r \geq 0} \tilde{p}_\tau \otimes_n H_n^{(r,n)}(t, x, y); \; H_n(t, x, y) = (L_n - \tilde{L}_n^*) \tilde{p}_\alpha(t, x, y), \end{aligned}$$

where $\tilde{p}_\tau(t, x, y) = \int_{\mathbb{R}^d} \tilde{p}_\alpha^x(\tau\Delta, x, z) \tilde{p}_\alpha^y(t - \tau\Delta, z, y) dz$ and L_n and $\tilde{L}_n^*$ stands for the discrete generators defined in equations (3.36) and (3.37). The reader may consult Konakov and Mammen [KM02] and Konakov and Menozzi [KM11] for more details.

We focus on the estimation of R_1. The arguments for R_2 are similar and hinted at the end of this proof. Observe that we can write:

$$R_1(1,x,y) = Q_M(1,x,y) + \frac{1}{n}\sum_{i=0}^{n-1}\int Q_M(t_i,x,z)\,\Phi(1-t_i,z,y)\,dz, \tag{3.58}$$

$$\Phi(1-t_i,z,y) = \sum_{r=1}^{+\infty} H^{(r,n)}(1-t_i,z,y). \tag{3.59}$$

Using Kolmogorov's forward and backward equations, one shows by induction on M that $Q_M(t,x,y)$ can be written as:

$$\begin{aligned} Q_M(t_k,x,y) &= \frac{1}{M!}\sum_{i=0}^{k-1}\int_{t_i}^{(i+1)/n}[n(u-t_i)]\int_0^1(1-\delta)^{M-1} \\ &\quad \int p(t_i+\delta(u-t_i),x,z)\,(L-\tilde{L}^*)^{M+1}\tilde{p}_\alpha(t_k-(t_i+\delta(u-t_i)),z,y)\,dz d\delta du. \end{aligned}$$

See e.g. equation (4.5) in [KM02] for the brownian case and equation (4.4) in [KM11] for the stable case. Thus, apart from the additional integrations w.r.t. δ and u, we see that Q_M is of the same nature that the terms we already dealt with. Therefore, adapting the above arguments allows us to derive α-stable estimates for this term. In fact, a precise study of Q_M allows us to derive that $\forall i \in [\![1,n]\!]$, $\forall \gamma \geq 0$, $\exists C > 0$ such that:

$$\left|\partial_y^\gamma Q_M(t_k,x,y)\right| \leq C t_k^{-\frac{\gamma}{\alpha}-(M+1)}\bar{p}_K^\alpha(t_k,x,y). \tag{3.60}$$

To get this bound, we actually show that it holds for

$$\bar{Q}_M(t_k,x,y) = \int p(t,x,z)\,(L-\tilde{L}^*)^{M+1}\tilde{p}_\alpha(t_k-t,z,y)\,dz,$$

independently of $t \in [0,1]$. The arguments differs depending if t is closer to 0 or t_k. In the first case (say, $t \leq t_k/2$), the singularities induced by taking the derivative along y in $\tilde{p}(t_k-t,z,y)$ are bounded by to $t_k^{-\frac{\gamma}{\alpha}-(M+1)}$, which is the announced singularity. When t is closer to t_k (say $t > t_k/2$), we transfer the operator $(L-\tilde{L}^*)^{M+1}$ on p by taking the adjoint, and we change variables to $z = y-u$. The singularities in $\partial_y^\gamma\left((L-\tilde{L}^*)^{M+1}\right)^T p(t,x,y-u)$ then yields the announced $t_k^{-\frac{\gamma}{\alpha}-(M+1)}$, and we conclude using the semi-group property of Proposition 3.3. Thus, for $t_k = 1$, we have $|\partial_y^\gamma Q_M(1,x,y)| \leq C\bar{p}_K^\alpha(1,x,y)$, which is the announced bound.

Now, for the second part of (3.58), we split the sum:

$$\begin{aligned} \frac{1}{n}\sum_{i=0}^{n-1}\int dz Q_M(t_i,x,z)\,\Phi(1-t_i,z,y) &= \frac{1}{n}\sum_{i\leq\frac{n}{2}}\int dz Q_M(t_i,x,z)\,\Phi(1-t_i,z,y) \\ &\quad + \frac{1}{n}\sum_{i\geq\frac{n}{2}}\int dz Q_M(t_i,x,z)\,\Phi(1-t_i,z,y) \\ &= S_1(1,x,y) + S_2(1,x,y). \end{aligned}$$

In $S_1(1,x,y)$, the time parameter is such that when differentiating along y, the singularities are negligible. Thus, we derive under the integral, and use bounds (3.60) and (3.45), to get a convolution of $\bar{p}_K^\alpha$ functions. For $S_2(1,x,y)$, we make use of the change of variable $z = y-u$, to get:

$$S_2(1,x,y) = \frac{1}{n}\sum_{i\geq \frac{n}{2}} \int duQ_M\left(t_i,x,y-u\right)\Phi\left(1-t_i,y-u,y\right).$$

In the stable case, when taking the derivative along y the bounds (3.60) and (3.46) yield:

$$\begin{aligned}
\left|\partial_y^\gamma S_2\left(1,x,y\right)\right| &= \sum_{\eta=0}^{\gamma} C_\eta^\gamma \frac{1}{n}\sum_{i\geq\frac{n}{2}}\int du\left|\partial_y^{\gamma-\eta}Q_M\left(t_i,x,y-u\right)\right|\left|\partial_y^\eta\Phi\left(1-t_i,y-u,y\right)\right| \\
&\leq C_\gamma\frac{1}{n}\sum_{i\geq\frac{n}{2}}\int du\bar{p}_K^\alpha(t_i,x,y-u)\left(1+\frac{1\wedge|u|}{1-t_i}\right)\bar{p}_K^\alpha(1-t_i,y-u,y) \\
&\leq C\bar{p}_K^\alpha(1,x,y)+\frac{1}{n}\sum_{i\geq\frac{n}{2}}\int du\bar{p}_K^\alpha(t_i,x,y-u)\left(\frac{1\wedge|u|}{1-t_i}\right)\bar{p}_K^\alpha(1-t_i,y-u,y).
\end{aligned}$$

Now, recall from definition of $\bar{p}_K^\alpha$, we have either $\bar{p}_K^\alpha(t_i,x,z)\leq C\bar{p}_K^\alpha(1,x,y)$ or $\frac{1}{1-t_i}\bar{p}_K^\alpha(1-t_i,z,y)\leq C\bar{p}_K^\alpha(1,x,y)$, so that:

$$\frac{1}{n}\sum_{i\geq\frac{n}{2}}\int du\bar{p}_K^\alpha(t_i,x,y-u)\left(\frac{1\wedge|u|}{1-t_i}\right)\bar{p}_K^\alpha(1-t_i,y-u,y)\leq\bar{p}_K^\alpha(1,x,y)\left(1+1\wedge|y-x|\right).$$

In the gaussian case, the proof is simpler, as the derivative of Φ is estimated by:

$$\left|\partial_y^\eta\Phi\left(1-t_i,y-u,y\right)\right|\leq\frac{1}{\sqrt{1-t_i}}\bar{p}_K^\alpha(1-t_i,y-u,y),$$

and we can directly conclude comparing the sum over η to a Beta function.

For R_2, we can take the derivative in y under the integral in the above expression to get:

$$\partial_y^\gamma R_2(t,x,y) = \frac{1}{(M+1)!}\int_0^1(1-\tau)^M\partial_y^\gamma\left[p^d\otimes_n\left(\tilde{L}_*-\tilde{L}^*\right)^{M+1}\tilde{p}_\tau^\Delta\right](t,x,y),$$

which is a term of the same nature as the second part of the expansion (3.38). In particular, one can show bounds on $\tilde{p}_\tau^\Delta(t,x,y)$, similar to those of Lemma 3.1. With these estimates at hand, we may use similar arguments to derive an α-stable estimate on R_2, for $\alpha\in(0,2]$. We leave the remaining details to the reader.

□

Remark 3.2. *The terms in the expansion* (3.38) *depends on* n*. As already pointed out in [KM11] and [KM02], it is possible to make this expansion independent of* n*, using the bounds on the difference between the usual time space convolution* $\otimes$ *and its discretization* $\otimes_n$*. For* $M=2$*, one derives the expansion:*

$$\begin{aligned}
\partial_y^\gamma(p-p_n)(1,x,y) &= \frac{1}{2n}\partial_y^\gamma\Big(p\otimes_n(L-\tilde{L}^*)^2p^d\Big)(1,x,y) \\
&\quad -\frac{1}{2n}\partial_y^\gamma\Big(p^d\otimes_n(\tilde{L}_*-\tilde{L}^*)^2p_n\Big)(1,x,y)+\frac{1}{n^2}\partial_y^\gamma R(x,y) \\
&= \frac{1}{2n}\partial_y^\gamma\Big(p\otimes(L-\tilde{L}^*)^2p\Big)(1,x,y) \\
&\quad -\frac{1}{2n}\partial_y^\gamma\Big(p\otimes(\tilde{L}_*-\tilde{L}^*)^2p\Big)(1,x,y)+\frac{1}{n^2}\partial_y^\gamma\tilde{R}(x,y) \\
&= \frac{1}{2n}\partial_y^\gamma\Big(p\otimes(L^2-(\tilde{L}^*)^2)p\Big)(1,x,y)+\frac{1}{n^2}\partial_y^\gamma\tilde{R}(x,y).
\end{aligned}$$

In the above expansion, $\partial_y^\gamma \tilde{R}(x,y)$ is a remainder term bounded by some stable density as $\partial_y^\gamma R(x,y)$.

Corollary 3.1. *Assume [**A**] holds. Recall m denotes the regularity of the coefficients of the SDE* (3.31)*. Let $M \in \mathbb{N}^*$, such that when $\alpha = 2$, $0 < M \leq m/2$, and when $\alpha < 2$, $0 < M \leq m - (d+4)$. Then, the following expansion holds:*

$$\theta^{*,n} - \theta^* = \frac{C_1}{n} + \cdots + \frac{C_p}{n^{M-1}} + o\left(\frac{1}{n^{M-1}}\right).$$

Proof. We prove that under the assumptions of Corollary 3.1, [**H-(M-1)**] holds. Let $M \leq m/2$ for $\alpha = 2$ and $M \leq m - (d+4)$ for $\alpha < 2$ and let $\gamma \in \mathbb{N}$ with $\gamma \leq M - 1$. From Theorem 3.3, under [**A**], expansion (3.38) holds up to order $M - 1$. Moreover, from Remark 3.2, this expansion can be made independent of n, namely

$$\partial_{y^d}^\gamma p(1,x,y) - \partial_{y^d}^\gamma p_n(1,x,y) = \sum_{k=1}^{M-\gamma-1} \frac{1}{n^k}\Gamma_k^\gamma(x,y) + \frac{\partial_y^\gamma \tilde{R}_n(x,y)}{n^{M-\gamma}}. \tag{3.61}$$

Thus, integrating equation (3.61) in the $d-1$ first variables yields for all $M \leq m/2$ for $\alpha = 2$ and $M \leq m-(d+4)$ for $\alpha < 2$ and $\gamma \leq M - 1$:

$$\partial_{y^d}^\gamma p^{X_1^d}(1,x,y^d) - \partial_{y^d}^\gamma p_n^{X_1^{n,d}}(1,x,y^d) = \sum_{k=1}^{M-\gamma-1} \frac{1}{n^k}\bar{\Gamma}_k^\gamma(x,y^d) + \frac{\partial_{y^d}^\gamma \bar{R}_n^\gamma(x,y^d)}{n^{M-\gamma}}. \tag{3.62}$$

where we denoted $p^{X_1^d}$, $p_n^{X_1^{n,d}}$ the marginal densities of X_1^d and $X_1^{n,d}$, and

$$\bar{\Gamma}_k^\gamma(x,y^d) = \int_{\mathbb{R}^d \times \cdots \times \mathbb{R}^d} \Gamma_k^\gamma(x,y) dy_1 \cdots dy_{d-1}.$$

The Gaussian bound on the remainder implies $\frac{\partial_{y_d}^l \bar{R}_n(x,y^d)}{n^{M-\gamma}} = \mathcal{O}\left(n^{-(M-\gamma)}\right)$, so that we have for all $M \leq m/2$ for $\alpha = 2$ and $M \leq m - (d+4)$ for $\alpha < 2$ and $\gamma \leq M - 1$:

$$\partial_{y^d}^\gamma p^{X_1^d}(1,x,y^d) - \partial_{y^d}^\gamma p_n^{X_1^{n,d}}(1,x,y^d) = \sum_{k=1}^{M-\gamma-1} \frac{1}{n^k}\bar{\Gamma}_k^\gamma(x,y^d) + o(\frac{1}{n^{M-\gamma-1}}). \tag{3.63}$$

Now, since $h(\theta) - h^n(\theta) = \mathbb{P}^x(X_1^d \leq \theta) - \mathbb{P}^x(X_1^{n,d} \leq \theta)$, taking $\gamma = 0$ in (3.63) yields:

$$h(\theta) - h^n(\theta) = \int_{-\infty}^\theta \left(p^{X_1^d}(1,x,y_d) - p_n^{X_1^{n,d}}(1,x,y_d)\right) dy_d = \sum_{k=1}^{M-1} \frac{\Lambda_k^0(\theta)}{n^k} + o(\frac{1}{n^{M-1}}),$$

where we denoted $\Lambda_k^0(\theta) = \int_{-\infty}^\theta \bar{\Gamma}_k^0(x,y_d)dy_d$. Thus, the first assumption in [**H-(M-1)**] holds.

We now turn to the expansion of the derivatives. From expansion (3.63) one easily gets:

$$\partial_\theta(h - h^n)(\theta) = (p^{X_1^d} - p_n^{X_1^{n,d}})(1,x,\theta) = \sum_{k=1}^{M-2} \frac{1}{n^k}\bar{\Gamma}_k^0(x,\theta) + o(\frac{1}{n^{M-2}})$$

and $\forall l \leq M-2, \forall (x,\theta) \in \mathbb{R}^d \times \mathbb{R}$

$$\partial_\theta^l h(\theta) - \partial_\theta^l h^n(\theta) \quad = \quad \partial_\theta^{l-1} p^{X_1^d}(1,x,\theta) - \partial_\theta^{l-1} p_n^{X_1^{n,d}}(1,x,\theta) = \sum_{k=1}^{M-l-1} \frac{1}{n^k}\Lambda_k^l(\theta) + o(\frac{1}{n^{M-1-l}})$$

where we denoted for consistency $\Lambda_k^l(\theta) = \bar{\Gamma}_k^{l-1}(x,\theta)$. Consequently, expansions (2.5) and (2.6) holds in **[H-(M-1)]**. It remains to check the local uniform convergence and the invertibility of $Dh(\theta^*)$. For the latter, recall that $Dh(\theta^*) = p^d(1,x,\theta^*)$. Also, we know that under **[A]**, stable bounds holds for $p(1,x,y)$ (see e.g. [Aro67] in the gaussian case, and [Kol00] for the stable case), that is $p(1,x,y) \asymp \bar{p}_K^\alpha(1,x,y)$. Thus, the left hand side of the previous inequality gives that $p(1,x,y)$, and *a fortiori* $p^{X^d}(1,x,y)$, is never equal to zero. Hence $Dh(\theta^*)$ is invertible. Finally the local uniform convergence is a consequence of expansion (3.38). □

4. Numerical illustration

To illustrate the method we consider a geometric Brownian motion $(X_t)_{t\in[0,T]}$ with dynamics given by

$$X_t = x_0 \exp((r-\sigma^2/2)T + \sigma\sqrt{T}W_t),\ t\in[0,T]$$

for which the quantile is explicitly known at any level $\ell \in (0,1)$. Indeed a simple computation shows that

$$\theta^* = x_0 \exp((r-\sigma^2/2)T + \sigma\sqrt{T}\phi^{-1}(\ell))$$

where ϕ is the distribution function of the standard normal distribution $\mathcal{N}(0,1)$. Let us note that the assumptions of Section 3 are not satisfied in this example and that nobody would devise any kind of Monte Carlo simulation in practice since the law of X_T is explicitly known for any time T. However the Black-Scholes model and its Euler scheme appears as a natural and often used benchmark to test and evaluate the performance of Monte Carlo methods. We use the following values for the parameters: $x_0 = 100$, $r = 0.05$, $\sigma = 0.4$, $T = 1$, $\ell = 0.7$. The reference Black-Scholes quantile is $\theta^* = 119.69$. We set $\gamma(p) = \gamma_0/p$ with $\gamma_0 = 60$.

Let us note that in order to implement the Richardson-Romberg stochastic approximation estimator we need to simulate discretization schemes of the Brownian diffusion with different steps $\Delta_r = T/nr$, $r = 1,\cdots,R$. We thus need to simulate *consistent Brownian increments* on intervals of the form $[(k-1)T/(rn), kT/(rn)]$, $r = 1,\cdots,R$. The coefficients to compute by induction the Brownian increments from small intervals up to the root interval of length T/n have been computed up to $R = 5$ for $\alpha = 1$ and up to $R = 3$ for $\alpha = 1/2$ in [Pag07], Section 5.

In order to illustrate the result of Theroem 2.1, we plot in Figure 1 the behaviors of $\sum_{r=1}^R \mathbf{w}_r\theta^{*,rn} - \theta^*$ for $R = 2,3,4$ and $n = 2,\cdots,15$. We estimate $\theta^{*,rn}$ by θ_M^{rn}, with $M = 10^6$ samples for $R = 2$ and $M = 10^8$ samples for $R = 3,4$ using the same Brownian motion for each R (see Remark 2.3). We clearly see that the Richardson-Romberg estimator efficiency increases with R and the method gives satisfying results with $R = 3,4$ for small values of n.

Let us observe that the asymptotic optimal parameters in Propositions 2.4 and 2.5 depend on the structural parameters: α, $|C_R|$, $C(\gamma,\underline{\lambda})$, $\mathbb{E}\left[|H(\theta^*,U)|^2\right]$, $|C_1|$. Let us note that in Proposition 2.2 and Theorem 2.1 one may show that the constants $|C_R|$ writes $C_R = Dh(\theta^*)^{-1}\tilde{C}_R$. Here one has $Dh(\theta^*) = p(1,x,\theta^*)/(1-\ell)$ so that $|C_R| = |\tilde{C}_R|(1-\ell)/p(1,x,\theta^*)$. We estimate $p(1,x,\theta^*)$ by $p_n(1,x,\theta_M^n) \approx (\mathbb{P}_x(X_1^{n,d} \le \theta_M^n + \varepsilon) - \mathbb{P}_x(X_1^{n,d} \le \theta_M^n - \varepsilon))/2\varepsilon$ which in turn is approximated by the crude Monte Carlo estimator $(2M)^{-1}\sum_{k=1}^M \mathbf{1}_{\left\{(X_1^{n,d})^k \le \theta^*+\varepsilon\right\}} - \mathbf{1}_{\left\{(X_1^{n,d})^k \le \theta^*-\varepsilon\right\}}$ with $M = 1000$, $n = 100$, $\varepsilon = 0.1$ leading to the value $Dh(\theta^*) = 2.56\times 10^{-2}$. Finally estimating $|\tilde{C}_1|$ for the crude SA estimator and $|\tilde{C}_R|$ for the Richardson-Romberg extrapolation method is a challenging task. Consequently we implement these methods in a blind way setting $|\tilde{C}_R| = 1$ for every R. We also set $\underline{\lambda} = Dh(\theta^*)$ and $\gamma_0 = 1/\underline{\lambda}$. Note that we have $\mathbb{E}\left[|H(\theta^*,U)|^2\right] = \mathbb{E}\left[|1-(1-\ell)^{-1}\mathbf{1}_{\left\{X_1^d \ge \theta^*\right\}}|^2\right] = \ell/(1-\ell)$. The optimal parameters for the Richardson-Romberg extrapolation method are set according to Proposition 2.4 namely

$$n(\varepsilon) = \left\lceil \left(\frac{2\alpha R}{\beta}+1\right)^{\frac{1}{\alpha R}} \mu_R^{\frac{1}{\alpha R}} \varepsilon^{-\frac{1}{\alpha R}} \right\rceil \quad \text{and} \quad M(\varepsilon) = \left\lceil \gamma_0^{\frac{1}{\beta}} \nu_R^{\frac{2}{\beta}} \left(1+\frac{\beta}{2\alpha R}\right)^{\frac{2}{\beta}} \varepsilon^{-\frac{2}{\beta}} \right\rceil.$$

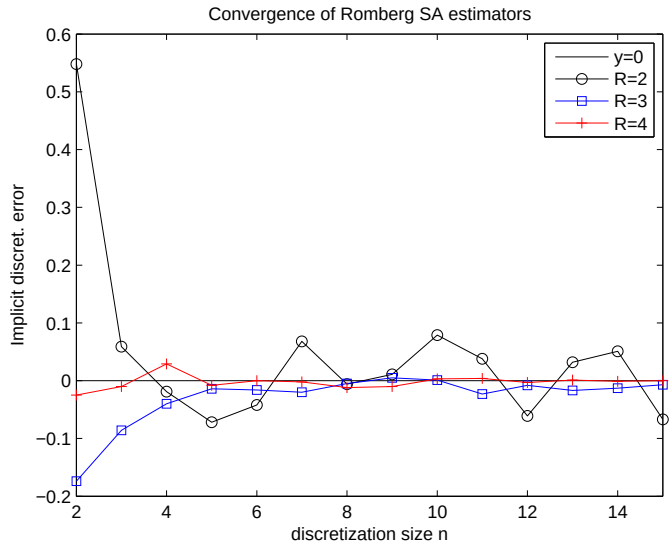

FIGURE 1. Richardson Romberg SA estimators: $\sum_{r=1}^{R} \mathbf{w}_r \theta^{*,rn} - \theta^*$ with respect to $n = 2, \cdots, 15$ for $R = 2, 3, 4$.

Target accuracy: ε	L^1-error	time (s)	R	n	M
5.00×10^{-1}	3.21×10^{-1}	0.9×10^{1}	2	14	8.69×10^{5}
2.50×10^{-1}	4.80×10^{-2}	5.15×10^{1}	2	20	3.48×10^{6}
1.25×10^{-1}	4.32×10^{-2}	1.70×10^{2}	3	8	1.21×10^{7}
6.25×10^{-2}	3.48×10^{-2}	7.92×10^{2}	3	10	4.85×10^{7}

TABLE 1. Richardson-Romberg SA estimators for the quantile at level ℓ of a geometric Brownian motion with a target accuracy $\varepsilon = 2^{-p}$, $p = 1, \cdots, 4$.

Target accuracy: ε	L^1-error	time (s)	n	M
5.00×10^{-1}	2.09×10^{-1}	$1,09 \times 10^{2}$	235	1.25×10^{6}
2.50×10^{-1}	3.84×10^{-2}	8.18×10^{2}	469	5.01×10^{6}
1.25×10^{-1}	3.48×10^{-2}	7.09×10^{3}	938	2.00×10^{7}
6.25×10^{-2}	2.91×10^{-2}	5.25×10^{4}	1876	8.01×10^{7}

TABLE 2. Crude SA estimators for the quantile at level ℓ of a geometric Brownian motion with a target accuracy $\varepsilon = 2^{-p}$, $p = 1, \cdots, 4$.

The target accuracy ε for the L^1-error has been set at $\varepsilon = 2^{-p}$, $p = 1, \cdots, 4$. The L^1-error is estimated using 400 runs of the algorithm. The results are summarized in Table 1 for the Richardson-Romberg extrapolation SA method and in Table 2 for the crude SA method.[1] Note that as expected the L^1-error is always lower than the specified ε for our estimators. Using the Richardson-Romberg SA scheme instead of the crude SA method leads to a gain in terms of CPU-time varying from 12 (for $\varepsilon = 5.00 \times 10^{-1}$) to 66 (for $\varepsilon = 6.25 \times 10^{-2}$).

[1]The computations were performed on a computer with 4 multithreaded(16) octo-core processors (Intel(R) Xeon(R) CPU E5-4620 @ 2.20GHz).

5. Technical results

We provide here some useful technical results that are used repeatedly throughout the paper. For a proof the reader may refer to [Fri13].

Lemma 5.1. *Let $a, b, \lambda > 0$. Let $(\gamma_n)_{n\geq 1}$ be a sequence satisfying* **(HS)**. *If $\gamma(t) = \gamma_0/t$, $t \geq 1$, suppose $b\lambda\gamma_0 > a$. Let $(v_n)_{n\geq 1}$ be a non-negative sequence. Then, for some positive constant $C := C(\lambda, \gamma)$, one has*

$$\limsup_n \gamma_n^{-a} \sum_{k=1}^{n} \gamma_k^{1+a} ||\Pi_{1,M}||^b v_k \leq C \limsup_n v_n,$$

where $\Pi_{k,M} := \prod_{j=k}^{M}(I_d - \gamma_j Dh(\theta^*))$, with the convention that $\Pi_{M+1,M} = I_d$.

Lemma 5.2. *Let $(\theta_p^n)_{p\geq 0}$ be the scheme defined by (1.3), θ_0^n being independent of the innovation with $\sup_{n\geq 1} \mathbb{E}|\theta_0^n|^2 < +\infty$. Suppose that* **(HUA)**, **(HC1)** *and* **(HS)** *hold. Then, for some constant $C > 0$, one has:*

$$\forall p \geq 1, \quad \sup_{n\geq 1} \mathbb{E}[|\theta_p^n - \theta^{*,n}|^2] \leq C\gamma(p)$$

References

[Aro67] D. G. Aronson. Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.*, 73:890–896, 1967.

[BMP90] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. Translated from the French by Stephen S. Wilson.

[Duf96] M. Duflo. *Algorithmes stochastiques*, volume 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996.

[FF13] M. Fathi and N. Frikha. Transport-entropy inequalities and deviation estimates for stochastic approximation schemes. *Electron. J. Probab.*, 18:no. 67, 1–36, 2013.

[FM12] N. Frikha and S. Menozzi. Concentration bounds for stochastic approximations. *Electron. Commun. Probab.*, 17:no. 47, 15, 2012.

[Fri64] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall, 1964.

[Fri13] N. Frikha. Multi-level stochastic approximation algorithms. *Preprint*, 2013.

[KM02] V. Konakov and E. Mammen. Edgeworth type expansions for euler schemes for stochastic differential equations. *Monte Carlo Methods Appl.*, 8–3:271–285, 2002.

[KM11] V. Konakov and S. Menozzi. Weak error for stable driven stochastic differential equations: expansion of the densities. *J. Theoret. Probab.*, 24(2):454–478, 2011.

[Kol00] V. Kolokoltsov. Symmetric stable laws and stable-like jump diffusions. *Proc. London Math. Soc.*, 80:725–768, 2000.

[KY03] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[LM10] V. Lemaire and S. Menozzi. On some non asyptotic bounds for the euler scheme. *Electronic Journal of Probability*, 15:1645–1681, 2010.

[Pag07] G. Pagès. Multi-step Richardson-Romberg extrapolation: remarks on variance control and complexity. *Monte Carlo Methods Appl.*, 13(1):37–70, 2007.

[RM51] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.

[Sat05] K. Sato. *Lévy processes and Infinitely divisible Distributions*. Cambridge University Press, 2005.

[She91] S. J. Sheu. Some estimates of the transition density of a nondegenerate diffusion Markov process. *Ann. Probab.*, 19–2:538–561, 1991.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. and App.*, 8-4:94–120, 1990.

[TZ04] D. Talay and Z. Zheng. Approximation of quantiles of components of diffusion processes. *Stochastic Processes and their Applications*, 109-1:23–46, 2004.